

Scientific and Technical Information Center

SEARCH REQUEST FORM

Date: _____ Requester's Full Name: _____ Examiner #: _____

Art Unit: _____ Phone (301) _____ Serial Number: _____

Results Format Preferred (circle): PAPER DISK E-MAIL

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Date: _____

Search Topic:

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known.

For Sequence Searches Only Please include all pertinent information (parent, grandchild, divisional, or issued patent numbers) along with the appropriate serial number.

FOR OFFICIAL USE ONLY

STAFF USE ONLY

Searcher: M. Smith 83278

Searcher Phone #: _____

Searcher Location: _____

Date Searcher Picked Up: 1-17-2000

Date Completed: 1-14-2000

Searcher Prep & Review Time: 15

Online Time: 15

Type of Search

2 NA Sequence (#)

1 AA Sequence (#)

Structure (#)

Bibliographic

Litigation

Fulltext

Other

Vendors and Cost

STN Dialog

Questel/Orbit Dr.Link

Lexis/Nexis Westlaw

WWW/Internet

In-house sequence systems (list)

Other (specify)

CompuGen

This Page Blank (uspto)

39	101	2.4	505	1	W26524	Eps11on-rhodomyces
----	-----	-----	-----	---	--------	--------------------

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

QY 373 GPVAPQENQSGSLGSPYHL--GPGRLRLVNNHRTSTPINNIIFGCIGRSESPDHYWIG 4300
| |::: | : ::||: | |||| :|| |||||::|

```

Dd	351	GD-CPSDKR---TDSTCHMTVSESANVLTAVSNVLKEIKILINIGVYIKGFEPDHYVVG	406
Qy	431	AQRDAMGGAAKSAVGTAILLELYRTSSMY-SNGFRRRLLEISMDGDFGVSSTEW	489
Dd	407	AQRDAMGGAARSGVGTALLKLQMFEVDMLKKGFPQRSIIIPASWSAGDFGVGATEW	466
Qy	490	LEGYSVYHLAAVYVSLDNNVLGDGDKFHAFTSPLLTSIESVLYKQVDSPPHSGQTLYEQ	549
Dd	467	LEGYSLSHLAAFTYINIDKRAVLTGSNFKVSASPLTLLIKTQNVKHP-VTGFQFLYQ-	523
Qy	550	VFTNPDSDAEYIRLPDSSAYSSTFAVGVPAVEFSEMEDDQAYPLHTEKEDYENLAK	609
Dd	525	---DSNW-ASKVEKLTLDNNAFPLAVSGIPAVSFCECDTD-YPLGTMTDYKLLIE	578
Qy	610	VLOGLP---AVAQAVQADQGLLIRLSHDLPLDGRYGVYVLRHGINENSGPLKA	666
Dd	579	---RIPLNKVVARRAAAEVQAFIKILHDELMLDERYNSQLTSFVRDLNGRAADKE	634
Qy	667	RLTLQWVYSARGDVIIRAELKROEYISSEERDERLTMYNVRIMRYEFLYSQYSPAD	726
Dd	635	MGLSIQWLYXSARGDFFRKRTSKLTIDFGNAEKTDFPVAKKLNDKRWARDEHLSYVSPKE	694
Qy	727	SPFRHIFENGDDTIGALLDLHLRLRSNSGCTPGATSTGPOESRRFRQALATLWTLOGA	786
Dd	695	SPFRHVFVSGSHLPALLEMLKLRKQNG-----AFNETLFRNQALATLWTLOGA	743
Qy	787	ANALSGDYMNIDNNF 801	
Dd	746	ANALSGDYWDIDNEF 760	
RESULT 2			
W22212	2		
ID	W22212	standard; Protein; 622 AA.	
AC	W22212:		
DT	02-NOV-1997	(first entry)	
DE		Rat transferrin receptor.	
KW		Radiation; radiometric; exposure; transferrin receptor:	
KW		red blood cell; rat.	
OS		Rattus norvegicus.	
PN	W09713630-A1.		
PD	01-MAY-1997.		
PF	21-OCT-1996; U16880.		
PR	24-OCT-1995; US-547197.		
PA	(UYNY) UNIV NEW YORK STATE RES FOUND.		
PI	Glomskl CA, Gong JK;		
DR	WPI; 97-259141/23.		
PT		Detecting exposure to radiation or radio-mimetic agents - by	
PR		correlation with a protein, such as transferrin receptor, present on	
PS		the surface of red blood cells of a mammal	
PS		Disclosure: Page 61-62; 83pp; English.	
CC		This polypeptide comprises the rat transferrin receptor (TR).	
CC		Claimed methods for detecting a mammal's prior exposure to	
CC		radiation or radiomimetic agents (RAs) involve: isolating red blood	
CC		cells from the mammal, and detecting the quantity of TRs on the red	
CC		blood cells using e.g., an antibody raised against the TR. The	
CC		quantity of TRs on the red blood cells is correlated with the	
CC		mammal's prior exposure. It is possible to assess accurately	
CC		cumulative lifetime exposure and to directly study the effects of	
CC		low doses of radiation, e.g. on cancer risk and other delayed	
CC		responses. The limited amount of blood necessary permits the test	
CC		to be conducted as frequently as desired. The test's sensitivity,	
CC		extending from 0.1 cGy to over 600 cGy, spans both the environmental	
CC		and the lethal dose ranges.	
QO		Sequence 622 AA;	

```

Query Match      33 98; Score 1424.5; DB: 1; Length 622;
Best Local Similarity 46.58; Pred. No. 9,2e-128;
Matches 304; Conservative 98; Mismatches 209; Indels 43; Gaps 15;

QY 155 DTRQTS---LERRVAGSGMALPQDIDIAALSRQKLDHWMTDFHYVGGQFPDPAPNTLL 211
| | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db	5	DIRKLSQNTTTPRAGSGQKQBNLAAYIENLFHDFKSKSWKRDHYHKKIQVKNVSQNL	63
Qy	212	HWVDEAGKVGESQDLPEDPDVYCPYSALGNVTGELVYAHYGPEDLDLIRANGVDPGRLL	273
Db	64	VYINSGSID---PVEAEGEYVAASKAGEYGLVTHANFGTKPKDFEEL---NYSVNGSLV	117
Qy	272	LVRGVISFAQKVTYNAOFGAQCGLIYEPADPESODPPKPSLSSQQAAYIGHVHLGTDPY	333
Db	118	IYRAKRIKFAEKVAAQSAFNAIGALYIDWRTF-----PVEDADLOFFGHAHGLTDPY	173
Qy	332	TPGFESFNQTOFPAPASSGLSPISAPQISADIASRLKELKGPAPADQMOSGLSGPYHL	393
Db	172	TPGFESFNHTQFPSPQSSGLSPISQVTSRAPKELKRNMGNL-CPSQSNID-SSKLEL	223
Qy	392	GGPPLRLVYNNHRTSTPINNIFCIEGRSEPDHYVYVIGAORDAMGPAQASAVGTALL	453
Db	230	SONQWVKTLLVNNVLTKETRIILNIFGIKYEEDPRYIVGAORDAMGPAQASAVGTALL	288
Qy	432	EYLVRFSSMYS-NGRRPKRSLFLISMGGDGGSGVSTHMLEGYLSVHLKAVVYVSLDNA	513
Db	290	KLAQFESMISMDGGRSPRSIIIFASMTAGDGAAGPPEMTEGYLSLHLKFTYINLDKY	349
Qy	511	VLGDDKFAKTSPLLTSLIEEYKQVDSPPNSGGTLEQVVFTRPSMDAEYIRPLPDSS	570
Db	350	VLGTSNFYKVASPPLTYTLMGKIMQDKRP-IDGKTYLR-----NSNMTSK-TEELSLDNA	407
Qy	571	AVSFAYEYGVPAVEESFEMEDDQAVPELTKTKEDYENLHKVLOGRLPAVAQV--AQLAG	627
Db	403	AFPELAYGIAVAVSCFCEDED-VYPLGTCKIDTYE---ILIQKVPOLQWVRYAAEVAG	457
Qy	628	QLLILSHDRLLPLDFGKYGQVYLRHIGNLNEFSGDLKARLTIQWYVSAGQDYYRAEK	688
Db	458	QFIKLTLDIELTDYEMYSKSLLSFEMDLNQFADIKDMLSLQWPLYSAQGDYFRATSR	517
Qy	688	LROEYISSEERDERLTRYNYNRIMKREVEFYLSQVYVSPADSPFRHIFMGCHTIGALLD	747
Db	518	LTTDHNMEKNNRPMARINDRIKAVHEHLSPIVSPSPESFRITFMGSGHTLSALVEN	577
Qy	748	LRLRNSNSSGTPGATSTGFOESRFRQLALLTWTLOGAANALSGDVMYNDNNE	801
Db	578	LRLRQKN-----ITAFNETLEFRQLALATLWTIQCVAANALSGDIWINDNEF	622
RESULT	3		
RS5097			
ID	RS5097 standard; Protein; 750 AA.		
AC	RS5097.		
DT	11-JAN-1995 (first entry)		
DE	Prostate-specific membrane antigen.		
KW	Prostate-specific membrane antigen; PSM; prostate cancer;		
KW	transmembrane glycoprotein; imaging; targeting; tumour detection;		
KW	antibody detection.		
OS	Homo sapiens.		
FT	Key	Location/Qualifiers	
FT	peptide	63..68	
FT	peptide	/note="region of high hydrophilicity 1,claim 35"	
FT	peptide	132..137	
FT	peptide	/note="region of high hydrophilicity 2,claim 35"	
FT	peptide	482..487	
FT	peptide	/note="region of high hydrophilicity 3,claim 35"	
FT	domain	23..39	
FT	domain	/note="specific membrane spanning domain"	
PN	W09409820-A.		
PD	11-MAY-1994.		
PF	05-NOV-1993; U10624.		
PR	05-NOV-1992; US-973337.		
PA	(SLOK) SLOAN KETTERING INST CANCER.		
PI	Fair WR, Heston MDW, Israeli RS;		
DR	WPI; 94-167129/20.		
DR	N-PSDB; 065520.		
PT	Prostate-specific membrane antigen and DNA encoding it - is		
PT	useful for detecting hematogenous micro-metastatic tumour cells.		
PT	and for identifying ligands which bind to PSM Ag		

PI Fair WR, Heston MDW, Israeli RS;
DR WPI: 94-167129/20.
DR N-PSDB; 065520.
PT Prostate-specific membrane antigen and DNA encoding it - is
PT useful for detecting haematogenous micro-metastatic tumour cells
PT and for identifying ligands which bind to PSM Ag

PS Claim 24: page 103-106: 196pp: English
CC R55097 is a prostate specific membrane antigen (PSM, encoded by
CC q05520). The inventors have identified 3 peptides of high
CC hydrophilicity to which antibodies may be directed against. In the
CC specification the specific membrane spanning domain is stated to be
CC from amino acid residue 19 to residue 44. The PSM coding sequence
CC is useful for suppressing or modulating the metastatic ability of
CC prostate tumour cells to grow, or for eliminating them. The protein
CC is useful to identify or purify ligands of the Ag. It is also an
CC attractive target for Ab-directed imaging and targeting of prostatic
CC tumour deposits.
CC Sequence 750 AA:

Query Match	17.2%;	Score 725;	DB 1;	Length 750;
Best Local Similarity	27.3%;	Pred. No. 1.6e-60;		
Matches 213; Conservative	142;	Mismatches 323;	Indels 102;	Gaps 22

QY	65	RPRPNLIPMAAARBRAPYLVLALILFPGALILGVAFRGSQAQGSUVLSEBQVY	124
Db	17	RPR-----WLCAG-----ALVLAGEFLLGL-----FGWITKSNENETNI	52
QY	125	EPDDLHOGRLYMSDLOAMFLOFEGEGRLEDTIROTSLREHVASGAMALITODIRALS	184
Db	53	TPK---HNMKAFIDELK-----ENIKKFLVNFQIPHLAGTEQNPQLAGIOSOM	101
QY	185	RQKLDHWTDTIHYGLOPPDPANHTLHWVDEACKVEOLPLEDP-----DYQCPY	235
Db	102	EFGLDSVELAHYDVLSTPNKTHPNYISIIEDENEIFNTSLFEPPEGENSDIYPPE	161
QY	236	SAT---GNVTELYEYAHNGREPDIQDI-RANGVPCVRLLYRNGVLSFQOKYTNADFG	291
Db	162	SAFSPQMPESDLYVYNARTEDFFKLERDKIKCSKIVARIARGVKRGKNNKMAQIAG	221
QY	292	AQGVLLYPPADFSODPEKPSLSSQOAVYGHVH-----GTGDPYPTGFPSENOQFP	344
Db	222	AKGVLYLSDPADYFAPGKSVPCGMNLPGGVQNGNLLNLNGADPLTPGYPAHEVAYRR	281
QY	345	PVASS-GLPSTPAPDISADIASRLRLKGVAP-QEMQSSLGSPYHLGPG-----P	395
Db	282	GIAENVGLPSPVHPPIGYDDAKLLEKKGGCAPDSSMRGS-LKVPYVNGPFGFNSTQ	340
QY	396	RLRLVNNHRTSPFINNFCGTEGRSEPDHVVVGAORDMGPAAKSAVGTALILETVR	455
Db	341	KVKHHIHTNVTYIINYIGTLRAVEPDRYVILGGHDSWFEGCIDPQSGAAVHEIVR	400
QY	456	TESSWVNSNGFPRRSLLEFISWDGDFGVSGETWLEGYLVLHKAVVYSLDANVIGD	515
Db	401	SFGTLKRGKMGPRRTILFASMDABEFLGSLTEWAENSRLQGRGAYIINADSSIGNY	460
QY	516	KFHAKTSPLLSLIESVLKQVDSFNH--SGQITLYQVVFNPMSDAEVIIRPMDSSAYS	573
Db	461	TLRYDCPRLMYSLVHNHLKELKSPDEGEGKSLTESWTKKSPSEFGMPRISKLGSD	520
QY	574	FTAF---VGVAPESEMEDQ-----AYPLRHKEDTIEYHLNVLOGRLPAAQAANAOL	625
Db	521	FEVFFORGLISGARARTKMEIKNESGYPLIYHVEYELVEKFXDPMF-KYHLIYAQV	579
QY	626	AGOLLIRLSHRLPLDFGARGDVALRHIGNLNEFS---GDLKARGILTOWYVSARGDY	681
Db	580	RGWVWFELANSIVLPFCRODRAVYLKRYADYIYSIKKHQEMKTYISVSDSLFSAYKNF	639
QY	682	IRAAEKLQEOIYSEEBDERLTRMYNVRIMKVEFYFLSQYSPADSPF-RHIFMGARDHT	740
Db	640	TEIASKESERLODDKSNPILVLRMMNQIMLEAFIDPLGLP-DRPFYRHVIVAYPSSHN	698
QY	741	LGALLDHLRLRUSSSGTPGATISSTGQESR-----FRQDALLTWTLOGAANALS	791
Db	699	KYA-----GESFPGIYDALDFIESKVDSKAMGEVRYOYAAATPQAAATTLIS	747

ID	standard; Protein; 750 AA
W02234	

DT	04-NOV-1996	(first entry)
DE	Prostate-specific membrane antigen.	
KW	Prostate-specific membrane antigen; PSM; promoter; prostate cancer;	
KW	metastasis; therapy; diagnosis.	
OS	Homo sapiens.	
FT	Key	
FT	Location/Qualifiers	
FT	domain	
FT	20..43	
FT	/label= Transmembrane_domain	
FT	modified_site	
FT	/label= N-glycosylation_site	
FT	modified_site	
FT	121	
FT	/label= N-glycosylation_site	
FT	modified_site	
FT	140	
FT	/label= N-glycosylation_site	
FT	modified_site	
FT	153	
FT	/label= N-glycosylation_site	
FT	modified_site	
FT	336	
FT	/label= N-glycosylation_site	
FT	modified_site	
FT	476	
FT	/label= N-glycosylation_site	
FT	modified_site	
FT	638	
FT	/label= N-glycosylation_site	
FT	WO9626272-A1.	
PN	29-AUG-1996.	
PD	23-FEB-1996; U02424.	
PF	24-FEB-1995; US-394152.	
PR	02-JUN-1995; US-470735.	
PR	02-JUN-1995; US-466381.	
PR	(SLOK) SLOAN KETTERING INST CANCER RES.	
PA	Fair WR, Heston WDM, Israeli RS.	
PI	WPI: 96-402365/40.	
DR	N-PSDB: T36785.	
DR	DNA encoding alternatively spliced prostate-specific membrane	
PT	antigen - useful to develop probe for detecting haematogenous	
PT	micrometastatic tumour cells, or prostate cancer progression	
PT	Example 1; Fig 47A-D: 284bp: English.	
FS	Prostate-specific membrane (PSM) antigen (W02234) is a type II	
CC	integral membrane glycoprotein that is highly expressed in prostatic	
CC	tumours and metastases and is almost entirely prostate-specific.	
CC	It provides an attractive cell surface epitope for antibody-directed	
CC	diagnostic imaging and cytotoxic targeting modalities. Its amino	
CC	acid sequence was deduced from a cDNA clone (T36785) obtd. from	
CC	human lymph node carcinoma of prostate cells. An alternatively	
CC	spliced variant of PSM (PSM') lacks the first 57 amino acids of PSM	
CC	and is probably cytosolic. Dctn. of PSM/PSM' tumour index can be	
CC	used to assess prostate cancer progression; PSM is the dominant	
CC	form in primary prostatic tumours.	
Q0	Sequence 750 AA:	

Query Match 17.2%; Score 725; DB 1; Length 750;
Best Local Similarity 27.3%; Pred. No. 1.6e-60;
Matches 213; Conservative 142; Mismatches 323; Indels 102; Gaps 22

OY	65	RPRQDNLPMMAACRRAPFLVTLFALLIFGAFLLGVAVRSGOACGSDVLVSEBQV	124
	17	RPR-----WLCAG-----ALVLAGFFFLGFL-----FGWFIKSNSENTNI	52
Db	125	EPDDLDFHOGRLVYSDLOAMFLQFGBEERLEDTIROTSLRERVAGSAGMAALTODIRALIS	184
OY	53	TPK---HMKKAFLELCKA-----ENIKKFLVNFQIPLHLAGTEGNPQLAIOISQWK	101
Db	185	RQKLDHVTDTTHYVGLQFPPDAHPNTHLWVDEAGKVGEOPLPDD-----DYICPY	235
OY	102	EFGLDSVELAHYDVLLSPKNTHPNYLSINBQNEJFNSTLEPPEPGYENVSADIPEP	161
Db	236	SAI---GNVTGELVYAHXGREDLODI--RARGDVPRGLLLVVRGVTSFAQKTNADQFG	291
OY	162	SAFSQGMPEBDLYVYNARLEDFEFLERDMKINCSCKIYIARYGKVFBNKKRKNQALAG	221
Db	292	AGQVLLYIPPAFDSQDPKPSLSSQOAVGVHNL-----GTGDPTYTGPCFSPFNQTOPF	344

```

Db 222 AKGVILYSDPADYFAPGVKSYDPGMNLPGGVQGRNLTNLNAGADPLTPGVAANYAYR 281
QY 345 PVASS-GIPSPAPISADIASRLRLKLGVPAP-QEMOGSLGSPYHLGCP-----P 395
Db 282 GIAEVNGLPISIPVHIGIYDDAKLEKMGSGSAPSSWMSG-LKPYNVGPGFTGNFSTQ 340
QY 396 RLRLVNNHRTSTPINNFCEGRSEPDHYVIGAGORDAMPAGAAKAVGATILLEYR 455
Db 341 KYKMHISTNEVTRITVYVIGIRGAVEPDRIYILGHRDQWVFGCIDPOSGAAVHEIVR 400
QY 456 TFSMVNSGFRPRRSLLFISWDGDFGSGVSTEMLEGYLSVHLKAVVYSLDNAVLGDD 515
Db 401 SFGTLKKGWRRRTIILFASMDAEFGLGSTEMAEBSRLQENGVAYINADSSIEGNY 460
QY 516 KFHATSPLLTSLIESVLKQVDSPNH--SGQTLIEQVYFTNPSMDAEVIRPLMDSSAYS 573
Db 461 TLRVDCPLMTSLVHNLTRELKSPDEGEGKSLYESWTKSPSPFGMPRIKSLGSDND 520
QY 574 FTAF---VGVPVAFESFEMDDQ---AYPLHTEKEDTYENLHKYLGRLPRAVAQAOL 625
Db 521 FEVFORLGISGARARYKNNMETNKFSGPLVHSVETELVEKYYDPMF-KYHLTVAQV 579
QY 626 AGQLLIRLSDRLPLDGRYGDVVLRLHIGNLNEFS---GDLKARGTLQWVYSARGDY 681
Db 580 RGVWFELANSIVLPFCRDYAVVLKRYADKIYISMKHPQMKYYSFDSLFSAYKNF 639
QY 682 IRAAKLKEIYISSEERDELTRMNVKRMVEEYFLSOYVSPADSPF-RHIFMGRCGHT 740
Db 640 TEIASKFESERLQDDKSNPIVLRMMNDQMLPRAFIIDPLGLP-DRPFYRHVYVAPSSHN 698
QY 741 LGALLDHLRLRNSSSGPGATSTGFOESR-----FRROLALLTWLQGAANLS 791
Db 699 KYA-----GQSPFGIYDALFDIESKVPDSKAMGEVKRQIYVAAFVQAATLTS 747

RESULT 5
ID W31524 standard; protein; 751 AA.
AC W31524;
DT 14-MAY-1998 (first entry)
DE Prostate-specific membrane antigen.
KW Prostate-specific membrane antigen; PSMA; PSM'; monoclonal antibody;
OS Homo sapiens.
FH Key Location/Qualifiers
FT 58..750
FT /label= "PSM"
FT /note= "Putative PSM' protein, a variant of PSMA"
FT Domain 20..43
FT /note= "Putative transmembrane domain"
FT Region 716..723
FT /note= "Peptide selected for monoclonal antibody development"

Mo9735616-A1.
02-OCT-1997.
PD 25-MAR-1997: U05214.
PR 25-MAR-1996: US-621399.
PA (PACT-) PACIFIC NORTHWEST CANCER FOUND.
PI Boynton AL, Holmes EH, Murphy GP, Tino WT;
DR WPJ; 97-489396/45.
PT Monoclonal antibody against prostate-specific membrane antigen -
PT used for diagnosis and treatment of prostatic cancer
PS Claim 2, Fig 1; 76pp: English.
CC The present sequence represents the prostate-specific membrane antigen
CC (PSMA). A novel monoclonal antibody (Mab) has been developed which has
CC an antigen-binding region specific for the extracellular domain (ECD) of
CC PSMA. The Mab is used for detecting the presence of PSMA (or its new
CC variant PSM', which lacks the intracellular domain (ICD)) in biological
CC specimens or cancer cells, particularly for the diagnosis, prognosis and
CC monitoring of prostate cancer. In which levels of PSMA are elevated.
CC Kits for carrying out this method are also included within the scope of
CC the invention. The Mab can also be used for treatment of prostatic

```

```

CC cancer and associated metastases, optionally when conjugated to a drug,
CC toxin or radioisotope. Mabs directed against ECD provide a more
CC effective detection than the known antibody 7E11-C15 which binds to the
CC ICD, and detects PSMA only in necrotic or apoptotic cells (and can not
CC detect PSM' at all). Tests for PSMA can be done on blood and urine,
CC without the need for a biopsy sample.
SO Sequence 751 AA;

Query Match 17.18; Score 720.5; DB 1; Length 751;
Best Local Similarity 27.28; Pred. No. 4.2e-60;
Matches 212; Conservative 142; Mismatches 325; Indels 101; Gaps 22;

QY 65 RPRQNLIPMAAARRAPYIVLALLIFGAFILGVAFRGSGOAGDSDLVYSEDEVY 124
Db 17 RPR-----WLCAG-----ALVLAGGFLLGFL-----PGWFKSSNEATNI 52
QY 125 EPDLDFHGRGLYWSLDQAMFLQFLGEGRLDPTIRQTSRERVAGSAGMAALTODIRALS 184
Db 53 TPK---HMKAFDLEL-----AENIKKFLINFTQIPHLAQTEDQNFOLAKOIOSOMK 102
QY 185 RQKLDHWTDTHYVGLQFPDPAPHPNTLHWDEAKVGEOLPLEDP-----DYVCPY 235
Db 103 EFGLDVSLAMVDVLLSYPNKTHPNYIISTINEDNEIFNTSLFEPPEGVENVDIYVPE 162
QY 236 SAI---GNVTELYYAHYGRPEDQLD-RANGVDPVGRLLVIRGVISFAQKVTNADDFG 291
Db 163 SAFSPQGMPEGLDYVNVARTKDEFFKLERDMKINCSKVIARAGKVRGNKRNADLAG 222
QY 292 AQGLVIEPEPADFDQDPKPKPLSSQQAIVYGVH-----GTGDPYTPGPPSPNQTFP 344
Db 223 AKGVILYSDPADYFAPGVKSYDPGMNLPGGVQGRNLTNLNAGADPLTPGVAANYAYR 282
QY 345 PVASS-GIPSPAPISADIASRLRLKLGVPAP-QEMOGSLGSPYHLGCP-----P 395
Db 283 GIAEVNGLPISIPVHIGIYDDAKLEKMGSGSAPSSWMSG-LKPYNVGPGFTGNFSTQ 341
QY 396 RLRLVNNHRTSTPINNFCEGRSEPDHYVIGAGORDAMPAGAAKAVGATILLEYR 455
Db 342 KYKMHISTNEVTRITVYVIGIRGAVEPDRIYILGHRDQWVFGCIDPOSGAAVHEIVR 401
QY 456 TFSMVNSGFRPRRSLLFISWDGDFGSGVSTEMLEGYLSVHLKAVVYSLDNAVLGDD 515
Db 402 SFGTLKKGWRRRTIILFASMDAEFGLGSTEMAEBSRLQENGVAYINADSSIEGNY 461
QY 516 KFHATSPLLTSLIESVLKQVDSPNH--SGQTLIEQVYFTNPSMDAEVIRPLMDSSAYS 573
Db 462 TLRVDCPLMTSLVHNLTRELKSPDEGEGKSLYESWTKSPSPFGMPRIKSLGSDND 521
QY 574 FTAF---VGVPVAFESFEMDDQ---AYPLHTEKEDTYENLHKYLGRLPRAVAQAOL 625
Db 522 FEVFORLGISGARARYKNNMETNKFSGPLVHSVETELVEKYYDPMF-KYHLTVAQV 579
QY 626 AGQLLIRLSDRLPLDGRYGDVVLRLHIGNLNEFS---GDLKARGTLQWVYSARGDY 681
Db 581 RGVWFELANSIVLPFCRDYAVVLKRYADKIYISMKHPQMKYYSFDSLFSAYKNF 639
QY 682 IRAAKLKEIYISSEERDELTRMNVKRMVEEYFLSOYVSPADSPF-RHIFMGRCGHT 740
Db 641 TEIASKFESERLQDDKSNPIVLRMMNDQMLPRAFIIDPLGLP-DRPFYRHVYVAPSSHN 698
QY 741 LGALLDHLRLRNSSSGPGATSTGFOESR-----FRROLALLTWLQGAANLS 791
Db 700 KYA-----GQSPFGIYDALFDIESKVPDSKAMGEVKRQIYVAAFVQAATLTS 748

RESULT 6
ID W47155 standard; Protein; 707 AA.
AC W47155;
DT 28-MAY-1998 (first entry)
DE Extracellular domain of prostate specific membrane antigen (PSMA).
KW Fc region; immunoglobulin; recombinant DNA; target protein; PSMA;

```

KM fusion protein; prostate specific membrane antigen; immunofusin;
 KM extracellular domain.
 OS Homo sapiens.
 PN US572604-A.
 PD 10-MAR-1998.
 PF 14-SEP-1995; 528122.
 PR 14-SEP-1995; US-528122.
 PR 14-SEP-1994; US-305700.
 PA (FUT-) FUT1 IMMUNOPHARMACEUTICALS CORP.
 PI Gillies SD, Lo K, Sudo Y;
 DR WPI: 98-192768/17.
 PT Recombinant DNA for expression of target protein, e.g. HIV gp120 -
 PT comprises sequences coding for signal peptide, immuno-globulin Fc
 PT region and gp120
 PS Example 14: Columns 27-32; 18pp; English.
 CC This is the extracellular domain of the prostate specific membrane
 CC antigen (PSMA) which is a type II membrane protein. The extracellular
 CC domain is modified to construct a new recombinant fusion protein, a
 CC PSMA immunofusin. The new recombinant DNA construct is used for
 CC expression and secretion of a target protein, whose sequence is free of
 CC immunoglobulin CH1 domain. The construct comprises a polynucleotide
 CC encoding from its 5' to 3' direction a secretion signal sequence,
 CC comprising a sequence encoding an immunoglobulin Fc region, and a
 CC sequence encoding the target protein. A replicable expression vector
 CC comprising the polynucleotide sequence can be used to transfect a host
 CC cell. The products can be used to produce a recombinant fusion protein
 CC (immunofusin) comprising the Fc region and target protein. The DNA can
 CC be expressed at high levels in a host cell, and the fusion protein is
 CC efficiently produced and secreted.
 SQ Sequence 707 AA;

Query Match 16.7%; Score 702.5; DB 1; Length 707;
 Best Local Similarity 28.4%; Pred. No. 2e-58;
 Matches 193; Conservative 128; Mismatches 291; Indels 67; Gaps 17;

QY 166 VASASAGMAALTODIRALSNQKIDHVTMDHYVGLQFPDPAHPTLMVDEAGKVGSDLP 225
 DB 40 LAGDEONFOLAKOIQOSWKKEGDLSDVELAHVDLSTPNKTHPYISITINDEGNEIFNTS 99
 QY 226 LEDP-----DYVCPSNAI--GNVTGELYVAHYGREDDLDL-RARGVDVGRLL 272
 DB 100 LFEPPREYGVNSDIYVFPFSAFSPQMPREGDLYVYNARTEDFKLERDKINSGKIV 159
 QY 273 VRVGVISFAQKVTINADFGAGVLIYEPADFSQDPKPSLSQAAVYGHVLT----- 326
 DB 160 ARRGKVRGNKKVKKNAQLAGAKGVILYSDPADYFAVPKVSYPDGNVLPGGQVGRNITLNLN 219
 QY 326 GTDDPTYPGPFSPFNOTQPPPAASS-GPSPAPDISADIASRLRLKLPAP-QEOMGS 383
 DB 220 GAGDPLTPGYPANETAYARRGAENVGLPSIPVHPHIGYDAOKLEKKGSAFPSSWRGS 279
 QY 384 ILGSPYILGG-----PRLRVNNHRTSTPINNIFGCIEGRSEPDHYVYIAQORDAW 436
 DB 280 -LKVPIYVGGEGFTGNSTQKAKHISTNEVTRYINVIIGIRGAVEPDRYITILGHRDSW 338
 QY 437 GPGAASAVGCTAILLELVRTSSMVSNGFRRRLSLFISMDGGDGVSGVSTEMLEGYLSV 496
 DB 339 VFGIDPQSGAAVYHEIVRSFGLTKKEGMRPRRTILFASMDAEFGGLSTEMAENSRRL 398
 QY 497 LHLKAVVYISLDNAVLDGDKFHAFTSPILSLISYVLKQVDSFNH--SGCITLQGVFTN 554
 DB 399 LOEGVAVYINADSSIEENYLLRVDCPLMWSLVNHLTKELKSPDEGEGKLSYKSWTKS 458
 QY 555 PSMDAEVIRPLPMDSSAYSFATF--VGVPAYEFSFEMDDO-----AYPFLHTKEDYEN 606
 DB 459 PSPFSGMPTRSKLSGNDVEVFQRLGASGRARYTKNMTNKPSPGLYHSHYETTEL 518
 QY 607 LHKVLOGRLPAVAQAVALQOLLIRLSHDLRLPLDFGRYGVYVLRHIGNLNEFS---G 662
 DB 519 VEKTYDPMF-KYHLTAQVRCGMVFEIANSIVLPDFCDRDVAVVRKYADKLYISMKHPQ 577
 QY 663 DLKARGLTLOWNVYARGDYIAAEKRLQELTSSSEDRERLTRMTNVRIMRVEFFLSQYV 722

DB 578 EMKTVSVSPDLSFSAKNFLEISKFSERLODDKSNPIYLRMMNQMLMELEAFIDPLG 637
 QY 723 SPADSEF-RHIFMGDHTLGLALDLRLRLNSSGTSGATSGTGFQESR-----F 772
 DB 638 LP-DREFYRHVIVYAPSSHNKYA-----GESFPGIYDALFDIESKVDPSKAMGEV 685
 QY 773 RQDALITWTLOGAANALS 791
 DB 686 KROIYVAAFVTOAAETLS 704

RESULT 7
 ID W33604 standard; Protein; 472 AA.
 AC W33604;
 DT 21-MAY-1998 (first entry)
 DE Human secreted protein AM282 full-length sequence.
 OS Secreted protein; AM282; cytokine; human.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..24
 FT Protein /label= Sig_peptide
 FT 25..472 /label= Mat_protein
 PN W09739030-A2.
 PD 23-OCT-1997.
 PF 16-APR-1997; 006475.
 PR 13-JAN-1997; US-783520.
 PR 18-APR-1996; US-634325.
 PA (GENEY) GENETICS INST INC.
 PI Jacobs K, Lavallie ER, McCoy JM, Werberg D, Racie LA,
 PI Spaulding V;
 DR WPI: 97-526400/48.
 DR N-PSDB: V02296.
 PT New isolated secretory proteins AM340, AM282 and AK583 - possibly
 PT have cytokine, cell proliferation/differentiation regulating,
 PT immunomodulating activities, etc.
 PS Claim 16; Page 45-47; 59pp; English.
 CC This human secreted protein, designated AM282, is encoded by a
 CC full-length cDNA clone (see V02296), deposited in ATCC 98026, that
 CC was identified from a database search using an isolated partial
 CC AM282 clone (see T97398). AM282 protein can be used in a claimed
 CC method for preventing, treating or ameliorating a medical
 CC condition. It may exhibit cytokine, cell proliferation (either
 CC inducing or inhibiting) or cell differentiation (either inducing or
 CC inhibiting) activity or may induce production of other cytokines in
 CC certain cell populations. It may also exhibit e.g. immune
 CC stimulating or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic or chemokinetic activity, haemostatic or thrombolytic
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC tumour inhibition activity, or other activities. No evidence of
 CC any of these activities is given in the specification.
 SQ Sequence 472 AA;

Query Match 3.3%; Score 140; DB 1; Length 472;
 Best Local Similarity 17.8%; Pred. No. 9.3e-05;
 Matches 106; Conservative 92; Mismatches 216; Indels 182; Gaps 25;

QY 93 FTGAFLIGYVAFRGSCQ-----ACGDSVLVSEDEVYEPDLDFHGRGLY 136
 DB 9 FCGVHLLSLCSGKAICXNGISKRTFEEIKELIASCQVAKAIIILVAYGK---AQNRSY 64
 QY 137 WSDQAMFLOFLGGRLEDIIRQTSLEBRYVAGSAGMALNODIRAAISROKLDHVTMDTH 196
 DB 65 --ERLALLVTVG--*-----PRLSGSKNEKAIQIMONQODGLEAV-----H 104
 QY 197 YVGQFPDPAHPNLIHNVDEAGKVGEOLEPDDVPYCPY-----SAIG-----NVTGELYV 247
 DB 105 LEPVRIP-----HW--ERGE--ESAVMLEPRIHAIILIGLSSIGTPPEGITAEIV 152

```

QY 248 AHYGRPEDDQDARAGVDPVGRLLVAVGVISFAQV-----INADFGAGVLYIPE 300
DB 153 V-----TSFDELQRRASEGRKIVVYNODYINYSRTVOYRTGGAVERAAGALASLI-RS 206
QY 301 PADFSQDPPKPELSSQQAAYGVHVLGTGDPYTPCFPSFNQTFPPVASSGLPSIPAQIS 360
DB 207 VASFSTIYSPHTIOGYO-----DGVKIKITPACT 235
QY 361 ADIASRLRLKLRKVPARPOEMOGSLGSPYHLGPGPRLRVNNHRTSPINNIFSCIEGR 420
DB 236 VDAEMKSMH-----ASHGIKIVLIQKKMGAKTPTDTSFNIV-AEITGS 278
QY 421 SPPDHVYVYGAQRDAW--GPGAASAVGTAILELVRTSSVAVSNGFRRRSLLPTISMDG 478
DB 279 KPEOVYLVSGHLDSDWGGAMDDGGAFISWEALSLIKDI--GLRKKRLRLVLMWA 335
QY 479 GDFGVSGSTEM-----LEGYLSVLHLKAVYVSLDNLVLDGDKFHAKTSPILSLIES 531
DB 336 GGGGVGAVAGYQVQLHKVNSNYSLVMSDAGTFLPTGLQFTSEKRA-----IMGE 387
QY 532 VLKQVDSPNHSGQTLVEQVVFNPMSDAEIVIRPLM-----DSSAVSFAVGVPAVE 584
DB 388 VM-----SLLQPLNITQVLSHGEGDINFVWIAQGVPGA- 421
QY 585 EFMEDDQAYPELHKEDTYENLHKVLOGRLPAVQAQVLAQGLLIRLSHRLLP 640
DB 421 -SLDDLYIKYFFFH--HSHGDTMTVMDPKMNVAAMVAVSYVAVD--EEMLP 470

```

RESULT 8
R99416 standard; Protein: 393 AA.

```

ID R99416.
AC R99416.
DE 04-DEC-1996 (first entry)
DE Amino peptidase precursor of Aeromonas caviae, useful in food mfr.
KW Amino peptidase; hydrophobic amino acid removal; hydrolysis;
KW food manufacture; improved taste.
OS Aeromonas caviae T-64.
FH Key Location/Qualifiers
FT 121.393
FT protein
FT J08173168-A.
PN 09-JUL-1996.
PD 09-JUL-1996.
PE 26-DEC-1994; 336663.
PR 26-DEC-1994; JP-336663.
PA (NORU) NORINSUISANSHO SHOKUHN SOGO.
PA (ZENK-) ZENKOKU RAKUNOGYO KYODO KUMIAI RENGOKAI.
DR WPI: 96-365587/37.
DR N-PSDB: T41568.
PT DNA encoding aeromonas microorganism derived aminopeptidase - which
PT eliminates bitter tastes by decomposing peptide(s) rich in
PT hydrophobic amino acids produced by protein hydrolysis, useful in
PT food mfg.
PS Claim 1; Page 5-7; 12pp; Japanese.
CC R99416 is an amino peptidase precursor from Aeromonas caviae. The
CC peptidase is useful in the hydrolysis of peptides high in hydrophobic
CC amino acids, this is useful in food manufacture as such peptides have
CC a bitter taste. The aminopeptidase removes amino acids sequentially
CC from the N-terminus of a peptide at an optimum reaction temperature
CC of 50 deg. C. and a pH of between 8 and 10.
SO Sequence 393 AA.

```

Query Match 3.1%; Score 132; DB 1; Length 393;
Best Local Similarity 20.9%; Pred. No. 0.0004;
Matches 96; Conservative 51; Mismatches 150; Indels 162; Gaps 18;

```

QY 256 IQDLRARGVDPVGRLLVAVGVISFAQVYTNADFGAGVLYIPEPADSPDKPSLSS 315
DB 12 LSPLAQAQAEPPWITVGADSGVELKQVKARLAPLFSASAPQLQAVESELGTLISHLMH 71
QY 316 QQAVGVHVLGTGDPYTPCFPSFNQTFPPVASSGLPSIPAQISDIAS----- 366

```

```

DB 72 E-----GHRGCG-----GIYVHS-----TLADALQSM-AQIPSONLFSAPPLTQASV 113
QY 366 -RLLRKLRKVPARPOEMOGSLGSPYHLGPGPRLRVNNHRTST----- 409
DB 114 NRLT-----PYLDQGNIVGITSQLASMR-----NRYTTTGVQASDVAWAGOMQSL 159
QY 409 -----PINNIFGCIERSPPDHVYVYGAQRDA-----WGCGAA 441
DB 160 SATLPWASVSKVSKGYPPQGSVLTLLKGRSRYPDVVYVGLHLDSTAGSAPNSRTLAPAD 219
QY 442 KSAVGTAILELVRTSSVAVSNGFRRRSLLPTISMDGDFGSYSTEWLEGLYSLHLKA 501
DB 220 DDASIALTLTEVLAV--IAQGRQPERTLOFTGTAAEVGLRSKDLA----- 266
QY 502 VVYVSLDNLVLDGDKFHAKTSPILSLIESVLKQVDSPNHSGQTLVEQVVFNPMSDAEV 561
DB 266 -----TRYKANTRKVLAAAL-----OLDMTNYGSA--EDIVMTDYDQGL 304
QY 562 IRPLMDSASVFAVGVPAVESEFM-----DQAVPF----- 597
DB 305 TGYLAQLLDAY-----LPQIRGYDSCGYGCSGDHASMHWQYPAAMPFESRENDYNPK 357
QY 597 LHTKEDTYENLHKVLOGRLPAVQAQV--AQLAGOLLIRL 633
DB 358 IHTADT-----LQNSDPSNAHLKRAQLATSFATEM 389

```

RESULT 9
W85456 standard; Protein: 472 AA.

```

ID W85456.
AC W85456.
DE 25-FEB-1999 (first entry)
DE Secreted protein encoded by clone bu45.2.
DE Secreted protein; nutritional activity; immune stimulating; vaccine;
KW Suppressing activity; haematopoiesis regulating activity;
KW tissue growth activity; activin; inhibin activity; chemotaxis;
KW chemokine activity; haemostasis; thrombolytic activity; receptor;
KW ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
KW tumour inhibition; gene therapy.
OS Homo sapiens.
PN WO9842739-A2.
PD 01-OCT-1998.
PE 20-MAR-1998; US05653.
PR 19-MAR-1998; US-044466.
PR 21-MAR-1997; US-822167.
PA (GENY) GENETICS INST INC.
PA Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
PI WPI: 98-609890/51.
DR N-PSDB: V82779.
PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult brain, foetal kidney, placenta or adult
PT pineal gland cDNA libraries.
PS Claim 14; Page 70-72; 113pp; English.
CC The present sequence represents a secreted protein. The polynucleotide
CC and secreted protein are predicted to have biological activities which
CC would make them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is given.
CC suggested activities include nutritional activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, haematopoiesis regulating
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity (no data is
CC given in the specification to support these activities). The
CC polynucleotide is also stated to be useful for gene therapy.
SO Sequence 472 AA.

```

Query Match 3.1%; Score 132; DB 1; Length 472;
Best Local Similarity 17.6%; Pred. No. 0.00054;
Matches 105; Conservative 92; Mismatches 217; Indels 182; Gaps 25;

OY FTGAFILGVAFRSCQ-----ACGSVLVYSADVWEEDLDHFHGRLLX 136
 | : | : | :
Db 9 FCGVHLISLCSCKAKICKNGISKRTFEEIKEIEKIASCAVAIINLAYVK---AQRSTY 64
 | : | : | :
OY 137 WSDIQAMELOELAGGRLEDTITROTSLREERVASGAMALNODIPAAALSROKLHWMTDTH 196
 | : | : | :
Db 65 --ERLALLVDIVG-----PRLSGSNLEKAIOIMQNQLQGEEKV----H 104
 | : | : | :
OY 197 YVGLOFPDPAPHPNTLHWYDAEKVGEOLPLEDDPVYCPR-----SAIG---NVYGELY 247
 | : | : | :
Db 105 LEPVRIP-----HW--ERGE--ESAVMLEPRHKIALIKGLSSSIGTPEPGITAFLV 152
 | : | : | :
OY 248 AHYGERPEDLODLRARGDVPVGRILLLVANGVISFEAKV-----TNQDGAGAVLYIRE 300
 | : | : | :
Db 153 V-----TSFDLQRRASSARKKIVYNOPYINSTRTOYRFGAVEAKVGAALSLI-RS 206
 | : | : | :
OY 301 PADFSODEPKPSLSNQOAVVGHVHLGNDPYTPGFSPFNQTOPPPVASGLPSIPAPQPIS 360
 | : | : | :
Db 207 VASTSIYSPHGIDETO-----DVCPKIPTACTI 235
 | : | : | :
OY 361 ADIASRLRLKLGPAVEDPWEGWSLLGPSYLGPGRRLRVNNHRSTPINNIIFGCIEGR 420
 | : | : | :
Db 236 VEDEMMSRM-----ASHGIKVITYOKMKAKTYPDPIDSNTV-AETIGS 278
 | : | : | :
OY 421 SEPDIHYVIGAORAW--GPCAASAVGTAILLELVTFESSWNSNGFRPRSLLFTSWDG 478
 | : | : | :
Db 279 KYPEEVOVYVSCHLSDMDVGOGAMDGGGAFISWEALSLIKDL--GLRPRTLRYLMWA 335
 | : | : | :
OY 479 GDIFSVCSTEW-----LEGYLEVLHKANVVYSLDNANVLGDGXFAKSPLLTSLIES 531
 | : | : | :
Db 336 EEOGVGVAFORYQLHKKNINISYSLWMSDGATFLPTGLQFTGSEKARA-----IMEE 387
 | : | : | :
OY 532 VLKOVDSEPNSHSQTLTEOVVFNPSPMAEVIRPLPM-----DSSASYTAFCVPAVE 584
 | : | : | :
Db 388 VM-----SLLPQNLINIQVLSHGEGTDINNWIQAAGVPGA- 421
 | : | : | :
OY 585 FSEMDDOAAYFFLTKEEDYEENLKHVLOGRLPAVAQAVALAQOLLIRLSHDRIIP 640
 | : | : | :
Db 421 -SLLDLIKYYEFFH---HSHGDTJTVMDPKOMNAVAAMVAVSYVADM--EEMLP 470
 | : | : | :
RESULT 10
ID R27481
R27481
AC R27481;
PT 03-MAR-1993 (first entry)
DE RP-III residual protease.
KW mutation; vpr: cleaning; laundry; detergent; washing powder; apr:
KM npr: epr; bpr: rp-I; mpr: rp-II; spoa gene;
KW reduced sporulation dependent protease.
OS Bacillus subtilis.
PN MO9216642.A.
PD 01-OCT-1992.
PF 26-FEB-1992: U01598,
PR 19-MAR-1991: US-671376.
PA (OMNI-) OMNIGENE INC.
P1 Pero J, Rufo GA, Sloma A;
DR WP1: 92-349223/42.
N-PDB: Q29134.
PT Bacillus residual protease III (RP-III) gene - which can be
mutated for prodn. of heterologous polypeptide(s) or expressed to
produce protease
Claim 25; Fig 4; 28pp; English.
CC This sequence represents residual protease III (rp-III) from
Bacillus subtilis, and was decoded from the appropriate DNA as detailed
in Q29134. The protease is secreted by B. subtilis in large amounts, and
can cause proteolytic degradation of any protein produced by the
bacteria, ie. a recombinant product. By introducing mutations in the
protease gene heterologous polypeptides can be expressed which do not
cause lowered levels of proteolytic damage. Also RP-III protease can be
expressed for use in improving the cleaning activity of laundry powders
or for use in industrial processes.

	Query Match Similarity	3.1%	Score 130:	DB 1:	Length 806;
	Best Local Similarity	21.8%	Pred. No. 0.0021:		
	Matches 101;	Conservative 52;	Mismatches 163;	Indels 148;	Gaps 23
QY	81	AAPYLVLTALLLTGTAFLTGLVYAFRGSOCAGSYLVVSEDNV-EPDL-DFHOGRLYW	137		
DB	164	SAPYVIGA-----NDAMPDGLYF-----GKSIKKAIIIDTGYEYNNHPDKRNGGYKGY-	210		
QY	138	SDLAAMPLOFL-----GEGRLDPTINGTSLRRVRAGSA-----	171		
DB	210	-----DFVNDYDPKETPTGDRGEATBDGTHAGVAGVAAANGITIKVAPDATLTATRYK	261		
QY	171	---GMAALTODIRAALSROKLDHWMTDTHYVS--LQFPDPANPTLHWYDEAGKY-----	221		
DB	262	LPGGSGTTEENIAGVERAVODGADVMNLSLNSINNPMATSTALDWMSEGVAVTSN	321		
QY	221	-----GEOPLREDDVY-YCPYS-----AIGNT	242		
DB	322	GNSGNGMTVSGPRTSGREALISGATOLPINEAYVFGSISAKVMGYNKEDDYKALNNE	381		
QY	243	GELYAHYGRPELDLDRAGVDPVGRLLLYRVGYISFAOKTNNODEGAOGLTYPE-P	301		
DB	382	VELLVETVAGIGEAQDFE-----GKDLGKVAVYVGRGSIAYFDKADKNKACATIGMYYVNNLS	436		
QY	302	ADFSSDP-----KPLSSSQQAAYYGHNLTGDPRTTGFFSFNQTORPPVAASSLPST	354		
DB	437	GTEIANVQMSVPTIKLSTLEDGEKLVSAKAGE-----TKTTFKLTYSKAL--	483		
QY	355	PAQPIASIDIASRLKRLKLPVAPRQEMOGSLGSPYHLGGPRRLRLVNNHRTSTPINTPI	414		
DB	483	-GEQY-ADFSSR-----GPVM-DTW---MIKRPISAGVNIIVSTPIPHNDHDHRYG-Y	527		
QY	415	GCIEGRSEPDHY-----VIGQRAMRGBAKSAV-GTALLLE	452		
DB	528	GKQGTSMASPHIAGAVAVIAQAKRKSVEQIIKAALIMNTAVPLK	571		

Query Match	2.9%;	Score 120.5;	DB 1;	length 500;
Best Local Similarity	23.3%;	Pred. No. 0.0076;		

[illegible]

RESULT 12

Y04868

ID Y04868 standard; Protein; 527 AA.

AC Y04868;

DT 06-JUL-1999 (first entry)

DE Mycobacterium species protein sequence 19f.

KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe; hybridisation; detection; vaccine; immunisation; infection.

OS Mycobacterium sp.

PN W09909186-A2.

PD 25-FEB-1999.

PE 14-AUG-1998; F01813.

PR 11-SEP-1997; FR-011325.

RA 14-AUG-1997; FR-010404.

PA (INSP) INST PASTEUR.

PI Glacquel B, Ilm EM, Pellicic V, Portnoi D, Gognet de la Salmoniere Y, Guigueno A;

PI WPI: 99-181045/15.

NP N-PSDB; X34120.

PT Mycobacterial DNA vectors containing reporter constructs - for identifying coding or promoter sequences involved in infection-associated protein expression

PT Claim 32, Fig 19f, 309pp; French

CS Sequences Y04742-Y05000 and Y07201-Y07204 represent secreted proteins from various Mycobacterium species microorganisms. The encoding nucleotide sequences can be used as primers and probes for methods for detecting and identifying mycobacteria, especially belonging to the M. tuberculosis complex. The encoded proteins can be used in vaccines for immunisation against a bacterial or viral infection.

Sequence 527 AA;

Query Match 2.9%; Score 120.5; DB 1; length 527;
Best Local Similarity 23.3%; Pred. No. 0.0083;

	Matches	125;	Conservative	55;	Mismatches	166;	Indels	171;	Gaps	30;
QY	161	SLEERVASGAGNAALT--ODIRALAS-----	ROKLDHWTDTHVYGLQFPDPAPH	208						
Db	74	ALRNRTSDTAMMAHLSTLQDIDIANANDGTGRTPGYQASVDVYVNTLRNSGFDVQTPERS	133							
QY	209	NLLHWVDE-----AKVGEQLPLE-----	DPDYVCYPSAIGVTELELYAHG-----	RP	253					
Db	134	ARVFKAEGGVTLTGNTVEARALEVSLGTPPD-----	GVTGPPLVAAPADSDPGSCP	184						
QY	254	EPLQDLRRAGVDPV--GRLLLVYGVGISFAOKVTNAQDRCAGCVLITPEPADRSQSPPKRS	312							
Db	185	SPYDL-----PYSGAVVYLDYRGVCPFRQKDDAAQAGAVALLI-----	AD-----	226						
QY	313	LSQQAVYVGHVHLGTDPEYTPGFPSPNOTQPPVASSGLPSPAPQISADIASRLRLKX	372							
Db	226	NIDEGAMGTGICANDT-----	YKLPVVSVTKSVGFQ--LRQGS	261						
QY	373	GPVAPQEMOGSLISGPYHLGPGPRLLTVNNHNRISTPLINNIFFGCTIEGRSEPDHYVIGAQ	432							
Db	262	GPTTVK-----LTAOSTP-----	KARNYIAQTKTGSSAN-----	VYAGAH	298					
QY	433	RBA--WGPRAKASNGATLLELVYTF--SSVYNSGFRPRRSLLTISMDGDFGSGVSGTE	488							
Db	299	LDVSVEPGEINDNGSGVAAVLETAVALGNSPRVSAVNR-----	FAFWGAEEFFIGISGN	352						
QY	489	WLEGYLSVYLHKAV--VYYSIDNAVY-----	GDDK--FHAKTSPLL--TSLIE--	531						
Db	353	YVES--LDIDALKGIALYLNFTDMLASPNRCYFTYDDQSLPLDARQOPVYPBSASGIERF	411							
QY	531	-SVLKQVDSPNHSGQTLLEYQVVFITNPMSDAEYIRPLPMDSSAYSTFAVGPAYVE--FSFM	588							
Db	412	VAYLTKM-----AGKTAQD-----	TSFGR-----	SDYDEFTLAGIRPSGGLFSGA	450					
QY	589	E-----DDQAYPLFTKREDYTNLHKVLOGRPLA--VAQAVQALQAL	629							
Db	451	EYKKSAGQAEALMGTADEPFDPENYHQKDDTLDIRTALGIMGAGVAAVAVGITYADLL	507							

RESULT 13
 ID W89614 standard; Protein; 496 AA.
 AC W89614;
 DT 17-MAR-1999 (first entry)
 DE Aspergillus oryzae dipeptidyl aminopeptidase #2.
 KW Prolyl dipeptidyl aminopeptidase; protein hydrolysate; dough;
 KW flavour enhancer; palatability; mouthfeel; aroma; crust colour;
 KW baking; animal feed additive; hydrolysis.
 OS Aspergillus oryzae.
 PN MO9851803-A1.
 PD 19-NOV-1998.
 PF 12-MAY-1998; U09629
 PR 20-OCT-1997; US-062892.
 PR 16-MAY-1997; US-857884.
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PI Blinkovsky A, Brown K, Byun T, Klotz A, Rey MW;
 DR WPI; 99-045232/04.
 DR N-PSDB: X00070.
 PT New dipeptidyl aminopeptidase from Aspergillus oryzae - used to
 PT produce protein hydrolysates enriched in particular amino acids,
 PT useful as flavour enhancers, e.g. in doughs
 PS Claim 1; Fig 1; 77pp; English.
 CC The present sequence represents dipeptidyl aminopeptidase (DPAP) from
 CC Aspergillus oryzae. DPAP acts synergistically with an aminopeptidase
 CC (AP) to hydrolyse polypeptides, producing protein hydrolysate (PH),
 CC useful in foods as flavour enhancer, e.g. in baked goods, enriched in:
 CC (a) Ala, Arg, Asp, Gly and/or Val, or (b), if the substrate has been
 CC demethylated, in Glu (free and/or peptide bound), in which case products
 CC are useful as animal feed additives. DPAP can also be used in flavour-
 CC improving compositions (optionally containing AP) and in dough pre-mixes
 CC also for deactivating enzymes and for converting precursors to mature
 CC proteins. DPAP increases the level of hydrolysates of proteins and thus
 CC of flavour development, and a mixture with AP may hydrolyse tripeptides

thus
peptides

CC that are resistant to either enzyme used alone. PH have improved
CC solubility, emulsifying and foaming properties, and products containing
CC them have better flavour, palatability, mouthfeel, aroma and crust
CC colour.
SQ Sequence 496 AA;

Query Match	2.8%	Score 117.5	DB 1	Length 496
Best Local Similarity	20.1%	Pred. No. 0.015		
Matches 108	Conservative 77	Mismatches 194	Indels 157	Gaps 28

QY	139	DLOAFNFIQFLGEG---	LEDITINQTSLEFVAGSAG---	MAALTDIRAA-----	LSRQK	188
Db	28	DIQ---	LEDLEGGSQLDEFA	FAIYERNRVFVGKRAH	DDTVNTLYIELKTKG	IVYKQDQ 84
QY	189	DHWITDTHYVGLQFPDPA	HPMTLHWADGAKVGEOLP	LEDDPDVCPYSAINVTGELY-		248
QY	248	-----AHYGRPELODL	RARGVDPVGRLLVYRGVLS	FAQKVTNMAODFGAGVLIPE		300
Db	123	KNLGCSEADY--PSDVE---	-----GKVALIKRGCEP	FGDKSVLAKAKAASIVYNN		170
QY	301	PADEFSQDPKPSLSSQA	AVYGHVHGTDPYPPGF	PFPEMOTQFPVASSGLEP	IPAQIS	360
Db	171	VA-----GSMAGTLGA	ASQDKG-PYS-----	ALYG-----IS		196
QY	361	ADIASRLIRKLK-LG	VPADPOEMOGSLGSPYH	LGPGRRLVYVNNHRTSTP	INNIFGCI	419
Db	197	LEDGQKTLIRKLEAG	SVSDVLA-----	-----VDSKOENRTTY	VVAQRTG	236
QY	420	RSEPHYVIVIGQRPAM	GPAAKSAVGTAILLELV	RTFSSMWSNCFRPRRSLL	FTSMQD	479
Db	237	-GDPNNVVALGGHTD	SVENGPEINDGSGITIS	NLV-----IAKALTOY	SVYNNARFLFTAE	292
QY	480	DFGSGVSTEMLEGYS	LVHLKAV-YVYSHD-----	NAVLA-----GDDKFAKTSPL	-----L	525
Db	293	EFGLLGSNYVY-SHL	NATLKNIRLYLNFDMA	SLASNYALMIYDGGSA	FNQSGPAGSAQI	351
QY	526	TSLSIVLKQVDSPN---	HSQCTLYEOWYFTNP	SPMDAEVIRPLPMDSSANS	FTAFVGP	581
Db	352	EKLFPDYDSDIDLP	PIPTQFDRSDYE-----	-AFILGIP-----SGGLFTAG	EGIM	398
QY	582	AVEFSEMEDDA-----	YPLHTRKEPTYENL-H	KVYLOGRLPAVAQAQIAGOL		629
Db	399	SEENASRWGGQAGV	AYDANY---HAAGDMTNL	NEAFLINSKATFAFAVATY	ANDL	451
RESULT 14						
ID	W89597	standard; Protein; 496 AA.				
AC	W89597;					
DT	17-MAR-1999	(first entry)				
DE	Aspergillus oryzae	aminopeptidase II.				
KW	Aminopeptidase; protein	hydrolyase; baking; dough; flavour enhancer;				
RV	mouthfeel; palatability;	aroma; hydrolysis; animal feed additive.				
OS	Aspergillus oryzae.					
PN	WP8851804-AL.					
PD	19-NOV-1998.					
PF	15-MAY-1998; U09940.					
PR	20-OCT-1997; US-062893.					
PR	16-MAY-1997; US-857886.					
PA	(NOVO) NOVO NORDISK	BIOTECH INC.				
PI	Blinskovsky A. Brown	K, Byun T, Golightly E, Kotod LV;				
DR	WPI: 99-045233/04.					
DR	N-PSDB: X00018.					
PT	New aminopeptidase	from Aspergillus oryzae - used to produce protein				
PT	hydrolyases enriched	in particular amino acids, used as flavour				
PT	enhancers, e.g. in	doughs				
PS	Claim 1; Fig 1; 92pp:	English.				
CC	The present sequence	represents aminopeptidase (AP) II from Aspergillus				
CC	oryzae. AP is used	in combination with an endopeptidase (EP) to				
CC	hydrolyse polypeptides,	producing protein hydrolyzate (PH), useful in				

CC foods a flavour enhancer, e.g. in baked goods, enriched in: (a) Leu,
CC Gly (especially), Glu, Ser, Asp, Asn, Pro, Cys, Ala and/or Gln, or (b),
CC if the substrate has been deamidated, in Glu (free and/or peptide bound)
CC in which case products are useful as animal feed additives. AP can be
CC used in flavour-improving compositions (optionally containing EP) and in
CC dough pre-mixes. Also AP can be used for deactivating enzymes and for
CC converting precursors to mature proteins. AP increases the level of
CC hydrolysis of proteins and thus of flavour development. PH have improved
CC solubility, emulsifying and foaming properties, and products containing
CC them have better flavour, palatability and aroma.

Query Match	2.8%	Score 117.5	DB 1	Length 496
Best Local Similarity	20.1%	Pred. No. 0.015		
Matches 108	Conservative 77	Mismatches 194	Indels 157	Gaps 28

```

OY 139 DQAMFLEOLFEG--RLDETITQOTSLREFRVAGSAC-----MALPQDIDRA-----LSROKL 188
Db 28 DIQ---LEDLEEGSQOLEDFAAIPERKNRVGGRKADHDVTANLYELKKTGYDYVKKQPO 84
OY 189 DHVWDTFHYVGLQFPDPAPHPTLTHWVDEAGKVEQOLPLEDDPVVCYPSAIGNVTGELYV- 248
Db 85 VHLMSN-----ADQTLKGVDE---EIEAKTMTSSSVTADVAVV 122
OY 248 -----AHYGRPEDLODLRAGVDPVGRLLTVRGVTSFAQKYNMAQDFGAGVLIPE 300
Db 123 KNLGCGSEADY--PSDVE-----GRVALIKRGECFPGDKRSVLAAKAKAASTAVINN 170
OY 301 PADFSQDPKPRLSPSSQAQAVYGHVILGTGDDPYTPGFPFSNQTOFPVAVSSGLPSIPAOPIS 360
Db 171 VA-----GSMAGTLGAQSDKG-PYS-----AIVG-----IS 196
OY 361 ADIARLLRLKRLK-GEVAQEQWQSGLLSPHYHLGPBRLRLVYNNHRTSPINNIGCIEG 419
Db 197 LEDGOKRLKLEAGSVSDVLM-----VDSKOENNTYVAVAQTG 236
OY 420 RSEPHYVYVIGAQRPAMGPAGAKSAGVTAILELTVFTSSMVSNGFPRPRLSLFTISMDCG 479
Db 227 -GDPNNVVALGSHTDSVEAGPEINDGSGIINLV--IAKALITQYSKNAVRFLEMTAE 292
OY 480 DFGVSQSTEWLEGYLSYHLKAV-VYVSLD-----NAVL-----GDDKFRAKTSPL---L 525
Db 293 EFGLLGSNYYV-SHINATELKNIRLYLNFDMIASPNVYLMIVDGGSAFNQSGPAGSAOI 351
OY 526 TSLISYVLKQVDSPN-----HSGOTLYEQVFTNPMSDADEVIRLPDMSASYSFTAFCVGP 581
Db 352 EKLFEFDYDSIDLPHIPIQFQDGRSDE-----AFILGIP---SGGLFTGAEGIM 398
OY 562 AVEFSFEMDDQA-----YPLFKTEDETYENL-HKVLOGRLPAAVAQAQVLAQOL 629
Db 399 SEENASRMVCGQAGVAVYDANY---HAAGDMNTMLNHEAFLINSKATAFAVATYANDL 451

RESULT 15
ID W89586
AC W89586 standard; Protein; 496 AA.
DI 17-MAR-1999 (first entry)
DE Aspergillus oryzae aminopeptidase II.
KW Aminopeptidase; protein hydrolyase; glycine releasing; protease;
KW proteolaceous material; flavour; food; baking; animal feed additive;
KW palatability; hydrolysis; solubility; emulsifying; foaming; aroma;
KW mouthfeel; crust colour.
OS Aspergillus oryzae.
PN WO9851163-A2.
PD 19-NOV-1998.
PF 15-MAY-1998. U09998.
PR 16-DEC-1997. US-069719.
PR 16-MAY-1997. US-857886.
PR 20-OCT-1997. US-062893.
PR 16-DEC-1997. DK-001465.
PA (ASAHI ) ASAHI CHEM IND CO LTD.

```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2000, 22:32:49 ; Search time 13.63 Seconds

(without alignments)
2772.019 Million cell updates/sec

Title: US-09-358-755-1

Perfect score: 4203
Sequence: 1 MERLKGFLFQRAQQLSPRSSO.....TLGGANALSGDVWNIDNPF 801

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database: PIR_62.*

Word size: 0

Number of hits that pass the threshold: 142080

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1504.5	35.8	757	2	A48592
2	1496.5	35.6	760	2	JXHU
3	1450.5	34.5	763	1	S29548
4	1428.5	34.0	776	1	JH0570
5	1424.5	33.9	622	2	A34549
6	725	17.2	750	2	A56881
7	438.5	10.4	811	2	S57149
8	297.5	7.1	783	2	S65188
9	243	5.8	809	2	S67153
10	208	4.9	63	2	S68317
11	182	4.3	455	2	S39663
12	143	3.4	504	2	S24314
13	136	3.2	561	2	E75191
14	131.5	3.1	565	2	E71223
15	130	3.1	806	2	A41341
16	123.5	2.9	495	1	P2M182
17	120.5	2.9	500	2	H70629
18	120	2.9	1448	2	T08526
19	120	2.9	1448	2	S37669
20	117.5	2.8	374	2	S69699
21	116.5	2.8	1301	2	S18118
22	116	2.8	1017	2	B70985
23	115.5	2.7	640	2	S62747
24	113.5	2.7	634	2	T00388
25	112.5	2.7	701	2	F48613
26	112.5	2.7	701	2	D48613
27	112.5	2.7	1603	2	A48613
28	112.5	2.7	1805	2	A44736
29	112	2.7	894	1	A41527
30	112	2.7	1128	2	T08312
31	111.5	2.7	1902	2	B44858
32	111	2.6	753	2	D72660
33	108.5	2.6	1167	1	A35066
34	108.5	2.6	1557	2	T13160
35	107.5	2.6	1902	2	B45764

35	107.5	2.6	695	2	D71283	probable translati
37	107	2.5	679	2	C71413	hypothetical prote
38	106.5	2.5	1902	2	S06997	serine proteinase
39	106.5	2.5	444	2	T01721	hypothetical prote
40	105.5	2.5	1257	1	A41060	neural cell adhesi
41	104.5	2.5	701	2	S35430	gag polyprotein -
42	104.5	2.5	2475	2	T00047	gelatin lyase (EC 4
43	104.5	2.5	1075	2	A57377	transcription fact
44	104	2.5	1330	2	B70836	hypothetical prote
45	104	2.5	1386	2	S73401	MG064 homolog R02_

ALIGNMENTS

RESULT 1
A48592
transferrin receptor protein - Chinese hamster
C:Species: Citicellus griseus (Chinese hamster)
C:Date: 03-May-1994 #sequence, revision 03-May-1994 #text, change 20-Aug-1999
C:Accession: A48592
R:COLIam, J.F.; Lai, A.; Domingo, D.; Fitch, M.; Hatton, S.; Trowbridge, I.S.
J. Biol. Chem. 268, 21686-21692, 1993
A:Title: YTRF is the conserved internalization signal of the transferrin receptor, an
A:Reference number: A48592; M01D:94012749
A:Accession: A48592
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-757 <COL>
A:Cross-references: GB:L19142; NID:g304528; PIDN:AAA03576.1; PID:g304529
C:Superfamily: transferrin receptor
C:Keywords: endocytosis; glycoprotein; transmembrane protein

Query Match 35.8%; Score 1504.5; DB 2; Length 757;

Best Local Similarity 42.8%; Pred. No. 1.5e-100; Matches 334; Conservative 119; Mismatches 248; Indels 79; Gaps 19;

Qy	35	EEEDDEGEAEALAHFCPELRGPEPLSRPNLIPMAAGRAAYLVLTALIFT	94
Db	44	EEENTDNMKASVVRKH---RLNGRLRCGT-----IAVVEFL	78
Qy	95	GAPLLGYAF--RGSCQAC-----GDSVLVVSSEVDNYNPPDLDFHOGRLYWSDLQAMF	144
Db	79	IGFMIGYGYCKRTEQKCVRLAETETGSEIIIOENIP-----QSSRLYMAIDLKXLL	131
Qy	145	LOFLGEGRLDTIROTSIREVAGSAGMAALTOIRALSNOKIDHWNTDHYVGLQFPD	204
Db	132	SEKIDALEFTITIKQLSOTSRDASQKDENLAYIENOFDFKLSKVRDEHYVKAIOVG	191
Qy	205	PAHPTLTHVDAGKVGQLEPLDEPDVYCPYSALGNVYGEIVYAHYGRPELDLQARV	264
Db	192	SAQNAVYIIVN---GSDLVENPGVYVASKATTVSGKLIHNFGRKRFEDLKY---	246
Qy	265	DPV-GRLLVAVGVISEAQVYTNADFGAGQVLIYEPADFSQDPKPSLSSQAAVGHV	323
Db	246	PVNGSLVIVRAGKITFEAKVANAQSFNAIGVLIYMDQTKF-----PVVEALESLFEGA	298
Qy	324	HLGSDPTPGFPSPFNQOPFPVASSGLPSTPAQDISADIASRLLRKAKGVAAQDMGGS	383
Db	299	HLGSDPTPGFPSPFNQOPFPVASSGLPSTPAQDISADIASRLLRKAKGVAAQDMGGS	357
Qy	384	LLGSFYLHGPPLRLVYNNHRTSTPINNIGCIEGRSEPDHYVYVIGAQDAMGGAAS	443
Db	358	SL-CKLESSGGINVLNVNNVKKERLITNITGVIKGFEPFRYIVYVGNQRAMGGAAS	416
Qy	444	AVGAILLELVITFSSWVS-NGFRRRSLLFISWDGDFGSGVSTEWLEGLVSLHLKAV	502
Db	417	SVGGLIKLAKAQAFAFSRSGFPRSRRIIFASWSAGDFGAVGATWLEGLVSLHLKAF	476
Qy	503	VYVSLDNAVILGDDKFAHKTSPILSTLESVLKQVDSPRHSQOTLYEQVVFNPMSDAVI	562
Db	477	TYINDKVVLTQTRNKFVSAAPLTYLLEIKTMQDVRRHP-IDSKPLR-----DSNMISKV-	530

QY 563 RPLPMSSAFTAFVGVPAVEFSFMDDOAYFFLTKEPTYENL-HKYLQGLRPAVAQA 621
D 530 EDLSLNAEPFLAYSGIPAVSWFCENED-YEYLDINLDYTEKLTOKVPO--LNNKVR 586
QY 622 VAOLAGOLRLRLSHDRLLPLDFGRYGDVVRHGINESFGDKARGLTLQWYSARGDY 681
D 587 AAFAVAGQFIKKLTHDIELNDYDMYNNKILSFYKELNQFADIKAGLSIQWLYSARGDF 646
QY 682 IRAAEKLRQETYSSEDERLRTMYNVRIMRVEFFYLSQVYSPADSEFRHIFMGQDHTL 741
D 647 FRATSLTDTDFHNAEKTRNFVREINNRIMKVEYHFLSPYSPRESEFRHIFMGQSHLT 706
QY 742 GALLDHLRLRSNSGPGATSTGQESFRROLALLTTTIGAAANALSGDWNINDNF 801
D 707 TALVENLKLKQNNSS-----AFNETLFRNQLALATWTIGQVANALSGDIDINDNF 757
RESULT 2
JXHU
transferrin receptor - human
N:Alternate names: CD7L; P90
M:Contains: 85k serum transferrin receptor
C:Species: Homo sapiens (man)
C:Date: 25-Feb-1985 #sequence.revision 25-Feb-1985 #text.change 22-Jun-1999
C:Accession: A93343; A90856; A36597; S54327; S09039; A03259
R:Schneider, C.; Owen, M.T.; Banville, D.; Williams, J.G.
Nature 311, 675-678, 1984
A:Title: Primary structure of human transferrin receptor deduced from the mRNA sequence.
A:Reference number: A93343; M0ID:85012743
A:Accession: A93343
A:Molecule type: mRNA
A:Residues: 1-760 <SCH>
A:Cross-references: GB:M1507; NID:q37432; PIDN:CAA25527.1; PID:q37433
R:McClelland, A.; Kuhn, L.C.; Ruddle, F.H.
Cell 39, 267-274, 1984
A:Title: The human transferrin receptor gene: genomic organization, and the complete pri
A:Reference number: A90856; M0ID:85048936
A:Accession: A90856
A:Molecule type: mRNA
A:Residues: 1-760 <MC>
A:Cross-references: GB:M1507; NID:q39915; PIDN:AAA61153.1; PID:q39916
R:Shih, Y.J.; Baynes, R.D.; Hudson, B.G.; Flowers, C.H.; Skikne, B.S.; Cook, J.D.
J. Biol. Chem. 265, 19077-19081, 1990
A:Title: Serum transferrin receptor is a truncated form of tissue receptor.
A:Reference number: A36597; M0ID:91035436
A:Accession: A36597
A:Molecule type: protein
A:Residues: 101-103, 'X', 105-108, 'X', 110-119 <SHI>
A:Experimental source: serum
R:Coppolino, M.; Migliorini, M.; Argaves, W.S.; Dedhar, S.
Biochem. J. 306, 129-134, 1995
A:Title: Identification of a novel form of the alpha(3) integrin subunit: covalent assoc
A:Reference number: S54327; M0ID:95169043
A:Accession: S54327
A:Molecule type: protein
A:Residues: 288-302;694-708;721-730 <COP>
R:Alvarez, E.; Girones, N.; Davis, R.J.
Biochem. J. 267, 31-35, 1990
A:Title: A point mutation in the cytoplasmic domain of the transferrin receptor inhibits
A:Reference number: S09039; M0ID:90226333
A:Accession: S09039
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19, 'C', 21-61 <ALV>
A:Note: mutant defective in endocytosis
C:Comment: This transmembrane glycoprotein exists as a dimer of similar or identical cha
ity acyl groups. The amino end of each chain lies within the cytoplasm and a stop-transf
embrane, may also serve as an internal signal sequence.
C:Comment: The expression of this receptor, involved in the regulation of cell growth, c
C:Genetics:
A:Gene: GDB:TFRC
A:Cross-references: GDB:120433; OMIM:190010

A:Map position: 3q26.2-3q26.2
C:Function:
A:Description: mediates cell iron uptake by binding, internalizing, and recycling the
C:Superfamily: transferrin receptor
C:Keywords: glycoprotein; iron transport; receptor; transmembrane protein
F:1-57/Domain: intracellular #status predicted <INT>
F:20-24/Region: tyrosine-based endosomal/lysosomal sorting signal
F:58-61/Region: stop-transfer sequence
F:62-89/Domain: transmembrane #status predicted <TM>
F:89-760/Domain: extracellular #status predicted <EXT>
F:101-760/Product: 85k serum transferrin receptor #status predicted <MAP>
F:251,317,727/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 35.6%; Score 1496.5; DB 1; Length 760;
Best Local Similarity 44.5%; Pred. No. 5; 7e-100;
Matches 327; Conservative 119; Mismatches 230; Indels 59; Gaps 20;

QY 84 YLVLTALLFTGAFLLGVAF-----RSCQAGDSVLVSEEDVYEPDLDPEHOG-RLY 136
D 68 YGTIAVVEFLIGFMIGYLGCKGVEKTECELAETSPVRE---EPGEPPAARRLY 123
QY 137 WSDLQAMFLQFLGEGRLDPTIR---QTSLERYAGSAGMAALTODIRALSRKLDHWT 193
D 124 WDLKRLLEKLDSTDTFTIKLNSYVPRHAGSQKXENALVYENQFREKLSKVMR 183
QY 194 DTHYVGLQPPDPAHPNTLWVDEAGKVGSQLPLEDDVYCPYSAIGNVTGELYAHYGRP 253
D 184 DQHFVLIQVKSQA--NSVIIVDKNGRL--VYLVENPGYVAASKATVGVKLVHANFCTR 240
QY 254 EDLQDIFRANGVDPV--GRLLVRYGVISFAOKYTNADDFAGOGVLIPEPADSQDPKPS 312
D 241 KDEEDL-----YTPVNGSIVYRAGKITFEAKVANAASLNAIGVLIYMDTKF-----PI 290
QY 313 LSSQAVYGVHVLGCDPYTPGPFSPFNQTFPPVASSGGLPSIPAPISADIASRLRLK 372
D 291 VNAELFEFGAHLGCDPYTPGPFSPFNQTFPPVASSGGLPSIPAPISADIASRLRLK 350
QY 373 GPVAPQEQGSLGSPYHL--GGPRLRLVNNHRTSTPINNIFGCIEGRSEPDHYVYG 430
D 351 GD-CPSPDK--TDSYCRVWTSKKNVKTIVSVLKEIKLINFYIKGFVEPDHYVVG 406
QY 431 AORDAMPGAASAVGTATLLELVRFSSMV--SNGRPRRLSILFIMDGDGSGVSTEM 489
D 407 AORDAMPGAASAVGTATLLELVRFSSMV--SNGRPRRLSILFIMDGDGSGVSTEM 466
QY 490 LECYLSVLHLKAVVYVSLDNVAVLGDDKFAKTSPLTSLIESVLYKQVDSPNHSGQTLYE 549
D 467 LECYLSVLHLKAVVYVSLDNVAVLGDDKFAKTSPLTSLIESVLYKQVDSPNHSGQTLYE 525
QY 550 VFTNSWDAEVYRPLPMSSAFTAFVGVPAVEFSFMDDOAYFFLTKEPTYENL 609
D 525 ---DSNW-ASKVEKLTLDNAEPFLAYSGIPAVSWFCENED-YEYLDINLDYTEKLE 578
QY 610 VLOGRLP---AQAQVLAQGLRLRLSHDRLLPLDFGRYGDVVRHGINESFGDKAR 666
D 579 ---RPELNAKARAALAEVAGQVYKLTIDVDELNDYENKNSLSEFYVDLNDYRADIK 634
QY 667 RGLTLQWYSARGDYIRAAEKLRQETYSSEDERLRTMYNVRIMRVEFFYLSQVYSPAD 726
D 635 MGLSLQWLYSARGDFRATSRLLTDTGNAEKTRNFVREINNRIMKVEYHFLSPYSPRE 694
QY 727 SPFRHIFMGQDHTLQALLDHLRLRSNSGPGATSTGQESFRROLALLTTTIGAAAN 786
D 695 SPFRHIFMGQDHTLQALLDHLRLRSNSGPGATSTGQESFRROLALLTTTIGAAAN 745
QY 787 ANALSGDWNINDNF 801
D 746 ANALSGDWNINDNF 760
RESULT 3
S29548

```

transferrin receptor - mouse
N:Alternate names: CD71; p90
N:Contains: 85k serum transferrin receptor
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S29548; A26735; D24550; 149662
R:Trowbridge, I.S.; Domingo, D.L.; Thomas, M.L.; Chain, A.
Submitted to the EMBL Data Library, January 1991
A:Reference number: S29548
A:Accession: S29548
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-763 <TRC>
R:Rothenberger, S.; Iacopetta, B.J.; Kuhn, L.C.
Cell 49, 423-431, 1987
A:Title: Endocytosis of the transferrin receptor requires the cytoplasmic domain but not
A:Reference number: A26735; MUID:87187639
A:Accession: A26735
A:Molecule type: Protein
A:Residues: 1-82 <ROT>
R:Grego, B.; Van Driel, I.R.; Stearne, P.A.; Goding, J.W.; Nice, E.C.; Simpson, R.J.
Eur. J. Biochem. 148, 485-491, 1985
A:Reference number: A24550; MUID:85203852
A:Accession: D24550
A:Molecule type: Protein
A:Residues: 7-19;158-175,'X',177-179;'DESC','AY',189,'TEN',193,'EXEF',195;196;197-208;45
A:Note: these trypsin fragments have been ordered by homology with the human sequence
R:Stearne, P.A.; Pietersz, G.A.; Goding, J.W.
J. Immunol. 134, 3474-3479, 1985
A:Title: cDNA cloning of the murine transferrin receptor: Sequence of trans-membrane and
A:Reference number: 149662; MUID:85159078
A:Accession: 149662
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 'AL',27-149,'Q',151-301 <RES>
A:Cross-references: GB:M29618; NID:g193272; PIDN:AA37616.1; PID:g193273
C:Comment: This transmembrane glycoprotein exists as a dimer of similar or identical cha
tracyl groups. The amino end of each chain lies within the cytoplasm and a stop-transf
embrane, may also serve as an internal signal sequence.
C:Function: The expression of this receptor, involved in the regulation of cell growth, c
C:Superfamily: mediates cell iron uptake by binding, internalizing, and recycling the ir
C:Keywords: glycoprotein; iron transport; receptor; transmembrane protein
F:1-57/Domain: Intracellular #status predicted <INT>
F:58-61/Region: stop-transfer sequence
F:62-89/Domain: transmembrane #status predicted <TMS>
F:89-763/Domain: extracellular #status predicted <EXT>
F:101-763/Product: 85k serum transferrin receptor #status predicted <MAT>
F:253,319,730/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match          34.5%  Score 1450.5; DB 1; Length 763;
Best Local Similarity 41.7%  Pred. No. 12e-96;
Matches 333; Conservative 127; Mismatches 257; Indels 81; Gaps 24;

```

```

Oy 24 ORVEGPRKGLHLE-----EEDEDEGAETLAHFCPELNGPPLGSRPPROPILIPMAAG 78
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 27 RQVDD-NSHVEMKLADEENADNMKA-----SVKPKPRFNR-----LCFAA-- 71
Oy 79 RRAAFLVLTALLITGAFLLGYA-FRGSQACGDS---VLVSEDEVNYPEDDFH-- 132
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 71 -----IALVIF---FLIGFMSOYLGYCKRVEQKECVKLAETFEETDKSETMETEDV 118
Oy 132 --QGRUYWSDLAQMFLOFEGEGRLEDTIROTS---LRERVAGSAGAAATQDIRAALSRO 186
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 119 PHSSTRLYADLTLSSEKLNSTIEFADTTIKQLSONYITTPREASQKDESLAYITENOFHEF 178
Oy 187 KLDHWYTDTHRYVGLQFPDPAHPTLHWVDEAKGVGEOLPLEDDPYVCPSYSAIGNVGLY 246
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 179 KRSKVWRBDEHYKIQVKSISGQNMVITVQSGNGLD---PVESPEGVVAHSKRTFVSGKLV 235

```

```

Oy 247 YAHVGRPEDLODLRARAGVPYGRLLTVRVGVISFAOKVTNNODEGACGLVYPEPADSQ 306
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 236 HANFTKKDFEEL---SYVNGSLTVIRAGETTFEAKVAAKNGSEVALVILYMDNKF-- 291
Oy 307 DPKPKSLSSOOAVYGVHVLGDTPTPGPFSPFNQOFPPVASSGSLPISAPISADIASR 366
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 291 ----PVLEADLALFHALGLGDPYTPFPFSPFNHQFPSSQSLPNIPIVQISAAAEK 346
Oy 367 LRLKIKFPVAPQEWGSLGSPRYLGPGRRLRVVNNHRTSTPIINNFGCLIEGRSEPDY 426
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 347 LFGKMEGS-CEARWID--SSCKLELSQONQVLYKVNILKERRIINIFGVIGYEEPRY 404
Oy 427 VVIGAQRDAMPBG-AAKSANGATILLETVRFESSVWS-NGFRPRSLFISMDGDFGV 484
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 405 VVVGQQRALGAVAKSVGGILLKLAOVFSDMISDGRPSHILFASWTAGDFGAV 464
Oy 485 GSTEMLECYLSVHLKAVVYSLDNAVGLGDDKFAKTSPLLTLSLESVLKQVDSNNHSGQ 544
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 465 GATEWLEGLSLHLKATTYINLDKVVGTSTFKYSASAPLTYLTMGKIMQDYKHP-VDGK 523
Oy 545 TLYEQVFTNPMDAEVIRPLPMDSSAYSTAFVGPVAFVSEFMDQAYPELATKEDTY 604
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 524 SLXR-----DSWISKV-EKLSFDNAVAPFLAYSGIPAVSFCCEDAD-YPLGTRLDY 576
Oy 605 ENL-HKVTQGRPAVAQVAGQLIRLSHDLPLPDEFGRYGQVYLRHIGNLNEFSGD 663
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 577 EALTQKVPQ--LNOVRRPAEYAGOLIKLTHDVLNLDYENSKLSTFMKDLNQFTKD 634
Oy 664 LKARLTLOWVYSGARDYIRAEKROEYSSSEEDERTRYANRIRKVEFFYSOYVS 723
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 635 IRDMSLSLQWLYSAGKDFRATSRITTFPHNAEKNNRFRINDRINKVEYHFHSPIYS 694
Oy 724 PADSPRRHIFMGDHTGALLDLRLSRNSSGTPGATSSGFOESRRROLALLTWL 783
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 695 PRESPRRHIFMGSGSHLSALVENIKLRKN-----ITAFNETLFRNQLATLATWTI 745
Oy 784 OGAAVALSGDVNIDNNF 801
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 746 OGVAVALSGDWINIDNEF 763

```

RESULT 4

```

transferrin receptor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: JH0570; S16855
R:Gerhardt, E.M.; Chan, L.N.L.; Jing, S.; Qi, M.; Trowbridge, I.S.
Gene 102, 249-254, 1991
A:Title: The cDNA sequence and primary structure of the chicken transferrin receptor.
A:Reference number: JH0570; MUID:91340160
A:Accession: JH0570
A:Molecule type: mRNA
A:Residues: 1-776 <GFR>
A:Cross-references: EMBL:X53348
A:Note: 581-His and 736-Gln were also found as the result of polymorphism
C:Comment: This protein mediates the endocytosis of the iron transferrin complex.
C:Superfamily: transferrin receptor
C:Keywords: glycoprotein; 11poptrotein; phosphoprotein; receptor; thiolester bond; tra
F:19-22/Region: coated-pit mediated internalization signal
F:70-88/Domain: transmembrane #status predicted <TRM>
F:23/Binding site: phosphate (Ser) (covalent) #status predicted
F:70/Binding site: palmitate (Cys) (covalent) #status predicted
F:261,326,391,738/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match          34.0%  Score 1428.5; DB 1; Length 776;
Best Local Similarity 40.0%  Pred. No. 4.7e-95;
Matches 324; Conservative 132; Mismatches 270; Indels 83; Gaps 20;

```

```

Oy 21 TYQRYVEGPRKGLHEE--EEDEDEGAETLAHFCPELNGPPLGSRPPROPILIPMAAGR 79
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 23 STAROTDGD-NSHVEMKLSADDEEG-----DIEPHEMHVSMAPQ-----RNCK 67

```


A;Residues: 66-149, 'P', 151-318, 'R', 320-414 <SID>

A₁Note: the authors claim that sequence of residues 41-57 is signal sequence
A₂Note: the authors translated the codon TTA for residue 34 as Phe, GAT for residue 201

C;Genetics:

C;Genetics:

A;Gene: SSP134

A;Map position: 16L

C; Superfamily: secretory protein SSP134

C;Keywords: glycoprotein

F:139,213/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.18; Score 297.5; DB 2; Length 783;

Best Local Similarity 22.48; Pred. No. 1.9e-13;

Matches 126; Conservative 107; Mismatches 258; Indels 71; Gaps 19;

```

0Y 152 RLEEDIRQTSIRERKAVSAGAMALTDODIRALRSQKIDHWTDHNYGLOFPRAHPNLT 211
Db 160 KIERLEIETISSMHPHSGTSGCALRHRHYIKESFDKNGIRLAGEEFMAVSNPGV---SL 216
0Y 212 HWMDAGKVGQBLPLEDDPVYCPYSAIGNVYG-ELVYAHYGREPELODDLARGVDVGYRL 270
Db 217 RYVSKDDTEGFDIPLEENF-FNPMSHNGQUNNIIVITANKASLUDMASQDQGLLNGDFI 275
0Y 271 LLVRGVTSFPAQKYTNADDFGAGVLLIYEPADFSGDPPKPSLSSQOAVYGVHVLGTDB 330
Db 276 LLVHGDVVFQOMLT-AQYEGAKAIIIFISEPYODNKD-----VIQKSVALLPQYGTGDA 328
0Y 331 YTPGPPSNQOPPPVASSGLPSIPADISMDISRLRLKLG---PVAQEWQSGSLGS 387
Db 329 LTPMEGSIROPIDATEKCLPKIPSTPISANQDKLTALISDTGVKFSNNLSSGSLNDC 388
0Y 388 PYHLGPRGLRLVYNNH-RTSTPFINNIFFGCIEGRESEPDHYVIGQADANGPAAKASAVG 446
Db 389 -----RLULLVQTAIREHPRPHDIYGRKEGSEQAGRALVIAPRANSAVYGTMPSTG 440
0Y 447 TAILLELVRTSSMVSN-GEPRRSLSEISMDGDFSSVSGSTMEGLYSVLHLKAVVY 505
Db 441 TIVVLLSTLIQLOEWMYKFKDMPLRNIFYISFGSEEFNAGATELMEKRTPEA-LKSEIYA 498
0Y 506 SLDNVAVLG--DD-KFAHKTSPLLSLLESLK-----QYDSNHSQGLIYEQVFTNP 555
Db 499 IIDVQOIQIMDSNNLETIOCHPLLVLDLOKNNKSRKFNKVDNVHOG-----547
0Y 556 SMDAEVIRPLEMDSASVFTAFVGVPAVEFSFMDDOAYPELHFKEDPTYENLHKVLQGR 615
Db 547 DMTPIYLAGIYV-----AIISSPGM-----NENH-ITYVEKPKPFIDKLRDOK 521
0Y 616 PA-VAQAAVQAQOLLIRLSHDLRLPLDFGRYGDVVLRLHIGNLNEFSGDLKARGLLQW 674
Db 592 KGEVSEIMELIVKSELIDDPFIPIFSISNYVDFLSTLTKDLQKCCPD-----TVNFD 645
0Y 675 YSARBDYIRAAEKLROEIIYSSE 696
Db 646 EVFLGTITLMENTKLOFEKWXSE 667

```

RESULT 9

S67153

probable membrane protein YOR256c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein 05330

C;Species: *Saccharomyces cerevisiae*

C:\Date: 12-Jul-1996 #sequence revision 12-Jul-1996 #text change 24-Sep-1999

C/Accession: S

R:Jauniaux, J.C.: Poirey, R.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67143

A: Accession: S67153

A: Accession: 56/153

A: molecule type: DNA

A;Residues: 1-809 <JAU>
A:CTGCG-CAATGCGCG:EMBI.4775164, NTD:G1420E79, DTD:CA000479 1: DTD:G35G100, DTD:G1420E79

A; Cross-references: EMBL:Z/5164; N1

A; Experiment

C;Genetics:

C; Superfamily: secretory protein SSP134

C;Keywords: transmembrane protein

F;126-142/Domain: transmembrane #status predicted <TMM>

Query Match	Score	DB 2	Length
5.88;	243;	809;	

Best Local Similarity 21.18; Pred. No. 1.7e-09;

Matches 169; Conservative 103; Mismatches 299; Indels 230; Gaps 30;

QY	83	PXYVITALLITFGAFLGUYAFRG---	SCQACGSLVYSEVDNYE---	PDLEFHGRLY	136
Db	120	PFILRRFFYIIEMSFILAYVSSYLEFNEKASGGKGFESCHDILFEYAKKASVDL-----		174	
QY	137	MSDLOAMELOFLGGRLEEDTIRQTSLEKRYAGSAGMALQODIRAAALSROKRLHWTDTH		196	
Db	174	-----AKFERDLEYISNPHSGSTGDAIARIYIOESPDDNNGKLTVEMGY		219	
QY	197	YVGLOFPDPAPHPNTLHWYDEAKGVEODLPLE-DEPYCPSYSAIGNTG-ELYAHNGRPE		254	
Db	220	SVYSNYPENV---SISYYDNK---NEKHDELSKENPNPLSSNGKLSKVALIYGKGTYY		273	
QY	255	DLDDIRAGVDPVGR---LLLVRYGISFPQKVTYMANQADFGAGOLIVREP-----ADESOD		307	
Db	274	DLQHKKSDKTLEDGKDYVLLQYDKIVS---QOVLIAKFGAKAVIIFISEPYGENIDVQOS		331	
QY	308	PPKPSLSQAQVAVGHVLTGDPPTPGFPSPFNQTOPPPVASSGL-		352	
Db	332	KP-----VGLPQYSTGD-----ASGLNMDGSPVEEKDKHFKWR		363	
QY	352	-PSIPAPISADIASRLRLK-KRPVAPQEMOGSLSPYHLHGPRPLRYVN-----NH		404	
Db	364	QTHLPTPISTRQCKELLSRLSSGGVYVD-----GNSDRSNGKGDVLLIDVLOTNV		417	
QY	405	RSTPFINNIFCIGRSEPHDYVYVIGAQORDAMGCAKSAVATLLELVRTFSM-VSN		463	
Db	418	REKHETPIRYVICKIGREGSDKAIITLASRNSINGTTPYPMFGTALLSYQLOQEVYKRF		477	
QY	464	GFRPRSLLETSMDGDFGVSGETMELEGYLSVLHKAUVYVSLDNAYLG---DDKF-HA		519	
Db	478	GWKPLRNIFYEFISFGTEFENYAGSSELVEORLT-PLKDEIYSLIDISQLOIPAEKXENG		535	
QY	520	KTSPLTMSLIESVYKO-----VDSRPHSQOTLYEQOVFFNPSMDADEVLRPLP		566	
Db	536	KIRGELSIEHPLLKKFPNRNNAHGNEDISVANOYHIG-----DWTPLLANGIP		563	
QY	567	MDSAYSTFAVGVPAVEFSEMEDDQAYPLPHTKEDYENLHKVLOGRLPVAAQVAYOLA		626	
Db	584	VSIVISSDSTRNRDPT-----ETSEDKFEFVERKLEDE---QNOQSVYDDL		626	
QY	627	GOLL---IRLSHDKRLPLDGFGRIDVYVLRHIGLINEFSGDLKARGTLQWVYSARGDYIR		663	
Db	627	YLLHISIMELIDDPLEHFDIISY-----		650	
QY	684	AAEKLRQIYSSERDEFLTMYNVMIRARVEFYL-----SOYSPADSPERHI		732	
Db	650	-----VEDIDERLOQLLEQAYPEKLNFTSIINGLLFMWKIGSEMAWSMTOG-WENI		697	
QY	733	FMGRGD-----HTIGALDLHLRLSRNSSGTPGAT-----SSTGFOESNFR		774	
Db	698	VMSHDDGIEPSSLISINRTWKKLTINIGRATCSAGLPPNSFYKKNVLFPGTLLIQEDSKN		757	
QY	775	OLALLTWTLOGAANAASGDVW	795		
Db	758	GGANDFWTFPGVMDAIYDDDD	778		

RESULT 10

568317

transferin receptor - quinea pig (fragment)

C:Species: *Cavia porcellus* (guinea pig)

C:\Date: 15-Feb-1997 #sequence revision 13-Mar-1997 #text change 20-Aug-1999

C: Accession: S68317

B:Gosiewska. A.: Mahmoodian. F.: Peterkofsky. B.
c,succession: 30031,
c,succession: 30031,

Db 282 YHGQAFIPKVNLE- IYANINLMWGSSDRNSTIMLVRTPLSRFSLVGVLEFLKE 340
 QY 536 V--DSPHSGQTLVEQVFFNPSPDAEYIRLPDMDSSAYSTAVGVPAVEFSMEDDQA 593
 Db 341 VNAGKSFSSGSP-PSYKFEKPEY-----MGSDHDFNFYSIPGVPIW-----PDRY 389
 QY 594 YPFLHTEDEYENLHKVLOGRLPAVQAQVLAQGLLIRLSHDLPLDPRGVGVLRH 653
 Db 390 Y---HTSADTPE---KLSRLTSLIGRAVY-ATGLFTAKAEKEEVERVARG---FAMKY 438
 QY 654 IGNL-----NEESGDLKARGLTLLQWVYSGDYIRAAEK---LRQETYS 695
 Db 439 LGELSMQRKTEVAESLYMNGSRDGLGLNVGHELESNG-VWVKKEKGIISINGLEYKN 497
 QY 696 EERDERL-----TRMTNVRIMRVEYFLSQ 720
 Db 498 ESIRARLKEIVEEDRMIVHLH--EYLMLE 526
 RESULT 15
 A41341
 Microbial serine proteinase (EC 3.4.21.-), minor (vpr), precursor - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 17-Jul-1992 #sequence.revision 17-Jul-1992 #text.change 29-Sep-1999
 C:Accession: A41341, B41341, S39700, D69730
 R:Strom, A.; Rulo Jr., G.A.; Theriault, K.A.; Dwyer, M.; Wilson, S.W.; Pero, J.
 J. Bacteriol. 173, 6889-6895, 1991
 A:Title: Cloning and characterization of the gene for an additional extracellular serine
 A:Reference number: A41341; MUID:92041574
 A:Accession: A41341
 A:Molecule type: DNA
 A:Residues: 1806 <SLQ>
 A:Cross-references: GB:W76590; NID:g143819; PIDN:AAA22881.1; PID:g143820
 A:Accession: B41341
 A:Molecule type: protein
 A:Residues: 161-195 <SLZ>
 R:Glasner, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu,
 A.; Rapoport, G.; Danchin, A.
 Mol. Microbiol. 10, 371-384, 1993
 A:Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region fr
 A:Reference number: S39655; MUID:95020537
 A:Accession: S39700
 A:Molecule type: DNA
 A:Residues: 1806 <GLA>
 A:Cross-references: EMBL:X73124; NID:g413923; PIDN:CAA51601.1; PID:g580871
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caidwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Galled
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koelter, P.; Koningsstein, G.; Krogh, S.; Kunano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon
 A:Authors: Schleich, S.; Schreier, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serot
 Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zunststein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: D69730
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-806 <KUN>
 A:Cross-references: GB:Z99123; GB:AL009126; NID:g2636240; PIDN:CAB15835.1; PID:e1186308;
 A:Experimental source: strain 168
 C:Comment: The amino terminal sequence of the mature protein and a molecular weight of c
 C:Genetics:
 A:Gene: vpr
 A:Start codon: TTG
 C:Superfamily: microbial serine proteinase vpr; subtilisin homology

C:Keywords: hydrolase; serine proteinase
 F:1-28/Domain: signal sequence #status predicted <SIG>
 F:28-160/Domain: propeptide #status predicted <PRO>
 F:180-548/Domain: subtilisin homology #status atypical <SBT>

Query Match 3.1%; Score 130; DB 2; Length 806;
 Best Local Similarity 21.8%; Pred. No. 0.24; Mismatches 163; Indels 148; Gaps 23;
 Matches 101; Conservative 52;

QY 81 AAPLYVTALLIFGAFLLGVAFRSGCAQGSVLVSEDEVNT-EPDL--DFHQGRLYW 137
 Db 164 SAPYIGA-----NDAMDLYT-----GKGIYALIIDGVENHNDLKKNFQYQYGY- 210
 QY 138 SDIQAMFLQFL-----GGRLEDYTRQTSLSRRVAGSA----- 171
 Db 210 -----DEVDNDYDPRKPTGPRGEANDHGHVAGTVAANGTIKGVAPDATLLAYRV 261
 QY 171 ---GMAALQDIRALSRKLDHWITDTHYV--LQPPDAHPTLLHWVDEAGV----- 221
 Db 262 LGPEGSGTTEWVLAGVERAVODGADVNNLSLGNLNNPDWATSTALDMAMSEGVVAVTSN 321
 QY 221 -----GEQLPLEDDVY-YCPYS-----AIGNVT 242
 Db 322 GNSGPNQWTVGSPQTSREAIISVGTQLPLNBYAVTFSGYSYSAKMYGKNKEDVYKALNKE 381
 QY 243 GELYVAHYGRPEDLDLRANGVDVGRLLLVYRGVVISFAQKVTNAODFGAOGVLIYEP-P 301
 Db 382 VELVEAGSIGCAKDFE-----GKDLTGKVAVVKRISIAFVDKADNAKAGAIQVYVNNLS 436
 QY 302 ADFEQDPP-----KPSLSQAQVYGHVHLGTDPTTPGPPSPNOQFPFVVASGGLPSI 354
 Db 437 GEIEANVPGMSVPTIKLSLEDGEXLSALKAGE-----TKTTFKLYYSKAL--- 483
 QY 355 PAQPIADIASRLRLKLGCPVAPQEMOGSLGSPYHLGPPRLRLVYNNHRTSTPINNIF 414
 Db 483 -GEYV-ADFSSR-----GPPM-DTW---MIKPDIAFPCVNIYSTIPTHDPHPYG-Y 527
 QY 415 GCIEGRSEPDHYV-----VIGAQRDAMPGANAASAV-GTALLLE 452
 Db 528 GSKQGTSMASPHIAGAVAVIKQAKPKMSVEQIKAAIMNTAVTTLK 571

Search completed: January 12, 2000, 22:33:58
 Job time: 69 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: January 13, 2000, 01:07:03 ; Search time 11.62 Seconds

(without alignments)
2004.979 Million cell updates/sec

Title: US-09-358-755-1

Perfect score: 4203
Sequence: 1 MERLMGLFQRAAQLSPRSSQ.....TLOGAANALSGDWNIDNMF 801

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database: SwissProt_38:*

Word size: 0

Number of hits that pass the threshold: 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1496.5	35.6	760	1 TRSR_HUMAN	P02786 homo sapien
2	725	17.2	750	1 PSM_HUMAN	O04609 homo sapien
3	438.5	10.4	811	1 YJ96_YEAST	P47161 saccharomyc
4	182	4.3	455	1 YWAD_BACSU	P25152 bacillus su
5	143	3.4	504	1 AMPX_VIBPR	Q01693 vibrio prot
6	130	3.1	806	1 SUBV_BACSU	P29141 bacillus su
7	123.5	2.9	467	1 VL2_BPV2	P06457 bovine papl
8	120	2.9	1448	1 TRC5_ECOLI	P27190 escherichia
9	112.5	2.7	1805	1 NEST_RAT	P21263 rattus norv
10	111.5	2.7	1902	1 P2P_LACPA	Q02470 lactobacilli
11	108.5	2.6	1167	1 SCPA_STRPY	P15926 streptococc
12	107.5	2.6	695	1 EFG2_TRBPA	O83748 treponema p
13	107	2.5	887	1 UFO_HUMAN	P30530 homo sapien
14	106.5	2.5	1902	1 P2P_LACLC	P16271 lactococcus
15	106.5	2.5	1902	1 P2P_LACLA	P15293 lactococcus
16	105.5	2.5	1257	1 CAML_HUMAN	P35004 homo sapien
17	104	2.5	519	1 AMYH_SACFI	P26989 saccharomyc
18	104	2.5	1386	1 Y064_MYCPN	P75613 mycoplasma
19	103	2.5	638	1 PAN1_RAT	P21676 rattus norv
20	103	2.5	930	1 PRGR_RABIT	P06186 oryctolagus
21	103	2.5	1295	1 PURL_ECOLI	P35254 escherichia
22	102	2.4	1319	1 ACSC_ACERY	P19450 acetobacter
23	100.5	2.4	2241	1 TEGU_HCMVA	P16785 human cytom
24	100	2.4	624	1 AMYG_ARYAD	P42042 axula aden
25	100	2.4	580	1 GAG_FIV	P10262 feline leuk
26	100	2.4	917	1 SLAP_THERH	P35830 thermus aqu
27	100	2.4	1170	1 YKD8_YEAST	P32862 saccharomyc
28	99.5	2.4	3567	1 ERY2_SACER	O03132 saccharomyc
29	99	2.4	632	1 RCAC_FREDI	Q01473 fremyella d
30	99	2.4	726	1 YANA_RHISN	P55577 rhizobium s
31	98.5	2.3	744	1 CA18_RABIT	P14282 oryctolagus
32	98.5	2.3	539	1 GAG_MLVFS	P26807 friend muri
33	98.5	2.3	1902	1 P3P_LACIC	P15292 lactococcus
34	98	2.3	519	1 AMYG_SACFI	P08017 saccharomyc
35	98	2.3	703	1 Y411_RHISN	P55492 rhizobium s
36	97.5	2.3	2594	1 7LES_DROYI	P20806 drosophila
37	97.5	2.3	868	1 AC02_SYNY3	P74582 synechocyst
38	97.5	2.3	1217	1 EGF_MOUSE	P01132 mus musculu
39	97	2.3	944	1 GYRA_SYNY3	P73077 synechocyst

ALIGNMENTS

RESULT	ID	TRSR_HUMAN	STANDARD:	PRT:	760 AA.	
40	97	2.3	631	1	PTBA_ERMCH	P26207 erwina chr
41	97	2.3	771	1	TLE3_MOUSE	O08122 mus musculu
42	96.5	2.3	701	1	GAG_RSV	P03322 rous sarcom
43	96	2.3	1666	1	CLH_SCHPO	Q10161 schizosacch
44	96	2.3	707	1	KPCT_MOUSE	Q02111 mus musculu
45	96	2.3	574	1	MP12_RAT	P48966 rattus norv
ALIGNMENTS						
RESULT	1	TRSR_HUMAN	STANDARD:	PRT:	760 AA.	
AC	P02786;	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)					
DT	15-JUL-1999 (Rel. 38, Last annotation update)					
DE	TRANSFERRIN RECEPTOR PROTEIN (TR) (ANTIGEN CD71) (T9) (P90).					
CN	TRFC.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
OC	Eutheria; Primates; Catarrhini; Homiidae; Homo.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE; 85012743.					
RA	SCHNEIDER C., OWEN M.J., BANYLLE D., WILLIAMS J.G.;					
RT	"Primary structure of human transferrin receptor deduced from the					
RT	mRNA sequence.";					
RL	Nature 311:675-678(1984).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE; 85048936.					
RA	MCCLELLAND A., KUHN L.C., RUDDLE F.H.;					
RT	"The human transferrin receptor gene: genomic organization, and the					
RT	complete primary structure of the receptor deduced from a cDNA					
RT	sequence.";					
RL	Cell 39:267-274(1984).					
RN	[3]					
RP	SEQUENCE OF 101-119 (SERUM FORM).					
RX	MEDLINE; 91035436.					
RA	SHIH Y.J., BAYNS R.D., HUDSON B.G., FLOWERS C.H., SKIKNE B.S.;					
RT	"Serum transferrin receptor is a truncated form of tissue receptor.";					
RT	J. Biol. Chem. 265:19077-19081(1990).					
RN	[4]					
RP	PALMITOYLATION OF CYS-62.					
RX	MEDLINE; 87218484.					
RA	JING S., TROMBRIDGE I.S.;					
RT	"Identification of the intermolecular disulfide bonds of the human					
RT	transferrin receptor and its lipid-attachment site.";					
RT	EMBO J. 6:327-331(1987).					
RN	[5]					
RP	MUTAGENESIS OF CYSTEINES INVOLVED IN INTERMOLECULAR BONDS.					
RX	MEDLINE; 90005427.					
RA	ALVAREZ E., GIRONES N., DAVIS R.J.;					
RT	"Intermolecular disulfide bonds are not required for the expression					
RT	of the dimeric state and functional activity of the transferrin					
RT	receptor.";					
RL	EMBO J. 8:2231-2240(1989).					
RN	[6]					
RP	INTERNALIZATION SEQUENCE, AND MUTAGENESIS OF TYR-20.					
RX	MEDLINE; 90130619.					
RA	JING S., SPENCER T., MILLER K., HOPKINS C., TROMBRIDGE I.S.;					
RT	"Role of the human transferrin receptor cytoplasmic domain in					
RT	endocytosis: localization of a specific signal sequence for					
RT	internalization.";					
RL	J. Cell Biol. 110:283-294(1990).					
RN	[7]					
RP	MUTAGENESIS.					
RX	MEDLINE; 94012749.					
RA	COLLAHN J.F., LAI A., DOMINGO D., FITCH M., HATTON S.;					
RT	TROMBRIDGE I.S.;					

RT	"YTRF is the conserved internalization signal of the transferrin receptor, and a second YTRF signal at position 31-34 enhances endocytosis."	
RL	J. Biol. Chem. 268:21686-21692(1993).	
RN	[8]	
RP	CARBOHYDRATE-BINDING SITES THR-104.	
RX	MEDLINE: 93043836.	
RA	DO S.I., CUMMINGS R.D.;	
RA	"Presence of O-linked oligosaccharide on a threonine residue in the human transferrin receptor."	
RT	glycobiology 2:345-353(1992).	
RN	[9]	
RP	CARBOHYDRATE-BINDING SITES THR-104.	
RX	MEDLINE: 93043837.	
RA	HAYES G.R., ENNS C.A., LUCAS J.J.;	
RT	"Identification of the O-linked glycosylation site of the human transferrin receptor."	
RL	glycobiology 2:355-359(1992).	
RN	[10]	
RP	STRUCTURE OF CARBOHYDRATE ASN-727.	
RX	MEDLINE: 95299226.	
RA	HAYES G.R., WILLIAMS A., COSTELLO C.E., ENNS C.A., LUCAS J.J.;	
RT	"The critical glycosylation site of human transferrin receptor contains a high-mannose oligosaccharide."	
RL	glycobiology 5:227-232(1995).	
CC	-1- FUNCTION: CELLULAR UPTAKE OF IRON OCCURS VIA RECEPTOR-MEDIATED ENDOCYTOSIS OF LIGAND-OCCUPIED TRANSFERRIN RECEPTOR.	
CC	-1- SUBUNIT: DIMER OF SIMILAR OR IDENTICAL CHAINS LINKED BY TWO DISULFIDE BONDS. A MONOMERIC SERUM FORM OF THE RECEPTOR EXISTS, IT LACKS RESIDUES 1-100.	
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.	
CC	-1- PTM: EACH CHAIN BINDS GLYCANS, PHOSPHATE, AND A PALMITATE GROUP.	
CC	-1- SIMILARITY: TO PROSTATE-SPECIFIC MEMBRANE ANTIGEN (PSM).	
CC	-1- DATABASE: NAME=PROV; NOTE=CD guide CD71 entry; WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd71.htm".	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or-send-an-email-to-license@sib-sib.ch).	
CC	-----	
DR	EMBL, X01060; CAA25527.1; -.	
DR	EMBL, M11507; AAA61153.1; -.	
DR	PIR, A03259; JXHU.	
DR	MIM, 190010; -.	
KM	Transmembrane; Glycoprotein; Receptor; Lipoprotein; Palmitate;	
KW	Signal-anchor; Endocytosis.	
FT	DOMAIN	1 62
FT	TRANSMEM	63 88
FT	DOMAIN	89 760
FT	SITE	58 61
FT	LIPID	62 62
FT	DISULFID	89 89
FT	DISULFID	98 98
FT	CARBOHYD	104 104
FT	CARBOHYD	251 251
FT	CARBOHYD	317 317
FT	CARBOHYD	727 727
FT	MUTAGEN	9 12
FT	MUTAGEN	20 20
FT	MUTAGEN	20 20
FT	MUTAGEN	20 23
FT	MUTAGEN	20 34
FT	MUTAGEN	20 34

[illegible]

```

RESULT 2
PSM_HUMAN STANDARD: PRT: 750 AA.
ID PSM_HUMAN 004609;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PROSTATE-SPECIFIC MEMBRANE ANTIGEN (PSM).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE: 93113576.
RA ISRAELI R.S., POWELL C.T., FAIR W.R., HESTON W.D.;
RT "Molecular cloning of a complementary DNA encoding a
RT prostate-specific membrane antigen."
RL Cancer Res. 53:227-230(1993).
CC - FUNCTION: MAY INTERACT WITH TRANSFERRIN OR ANOTHER LIGAND AND
CC POSSIBLY FACILITATES METASTATIC SPREAD.
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC - TISSUE SPECIFICITY: PROSTATIC EPITHELIAL CELLS.
CC - SIMILARITY: TO THE TRANSFERRIN RECEPTOR PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M99487; AAA60209.1; .
DR MIM: 600934; .
KW Antigen; Transmembrane; Glycoprotein; Signal-anchor.
FT DOMAIN 1 19 CYTOPLASMIC (POTENTIAL).
FT TRANSNM 20 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 44 750 EXTRACELLULAR (POTENTIAL).
FT MOD_RES 1 1 BLOCKED.
FT CARBOHYD 76 76 POTENTIAL.
FT CARBOHYD 121 121 POTENTIAL.
FT CARBOHYD 140 140 POTENTIAL.
FT CARBOHYD 153 153 POTENTIAL.
FT CARBOHYD 195 195 POTENTIAL.
FT CARBOHYD 336 336 POTENTIAL.
FT CARBOHYD 459 459 POTENTIAL.
FT CARBOHYD 476 476 POTENTIAL.
FT CARBOHYD 638 638 POTENTIAL.
FT SIMILAR 418 567 TO TRANSFERRIN RECEPTOR.
FT CONFLICT 354 354 R -> K (IN AA SEQUENCE).
SQ SEQUENCE 750 AA: 84330 MW: 73847D90 CRC32;

Query Match 17.2%; Score 725; DB 1; Length 750;
Best Local Similarity 27.3%; Pred. No. 5.5e-45;
Matches 213; Conservative 142; Mismatches 323; Indels 102; Gaps 22;

```

```

DB 162 SAFSPQMGPEGDLYVYVARTEDFEFLERDMKINCSGIYARIKGVGRGNKVNQAOLAG 221
QY AAGVLIYEPPEADFSQDPKPKSLSSQOAYGVHVL-----GTGPRTPGPFSEFQOTQFP 344
DB 222 AKGVLTYSBPDADYFAPGVKSYPDGWNLPDGGVQQRGNILNLNAGAGPLPGPYANETAYRR 281
QY PAAS-GLPSPAPQISADIASRLRLKLGYPAP-CEMOGSLSGPYHLGPG-----P 395
DB 282 GIAENAVGLPSIPVHPIGYYDAOKLEKMGGSAPRDSNRGS-LKPYVINGPQFTNFSQ 340
QY RLRLVYNNHRTSTPIINNIFGCIEGRSEPDHYVIGAQRDANGPGAASAVGATALLLEVR 455
DB 341 KKMHIISTNETVTRIVYVIGTLRGAVEPDRVYILGHDMSWFGIDIDPGSAAYVHEIYR 400
QY TFSSWVSGFPRRRLIISWDGDFGVSSTEMLEGLYSVLHLKAVVYSLDNAVLDGD 515
DB 401 SFGTLKKGWRPRRRIILPASMDAEFGILGSTWMEENSRLIOERGVAIYNADSIIEGNY 460
QY KFHATSPPLTSLIESVLYKOVDSPNH--SGOTIYQVVFNPMSDAEVIIRPLPMDSAYS 573
DB 461 TLRVDCPLMTSLVNLTKELKSPDEGEKSLYESWTKKSPSPFGMPRIKSLGEND 520
QY 574 FTAFA--VGVAPEESFEMDDQ-----APPLTKEDTYENLHKVLOGRLPAVAQAVAL 625
DB 521 FEVFFQRLGIASGRARVYKMETNKFSGYPLHSVYELVELKDYDMF-KYHLTVAQV 579
QY 626 AGQLIRLSDRLPLDFGRYGDVYLRIHGLNEFS---GDLKARGLTLOWVYASRDY 681
DB 580 RGMWFELANSIVLPDFORDYAVLRKYADKIYSISMHPQMKYYSFSLFSAVKNF 639
QY 682 ITRAKLQEIYSSERERELRMTNVMYIMVEFFLQIYSPADSP-RHIFMRGHT 740
DB 640 TEIASKESERLDDFKSNPIVLRMMNDQMLMEFRAIDPLGLP-DRPYRHYIYPSNN 698
QY 741 IGAALDHLRLKSNSSGTPGATSTGFQESR-----FRQALITWTLQGANALS 791
DB 699 KFA-----GESPGIYDALFDIESKDPKANGEVKROLYVAFTQAALITLS 747

RESULT 3
Y96_YEAST STANDARD: PRT: 811 AA.
ID Y96_YEAST 147161;
AC P47161;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 92.0 KD PROTEIN IN RPS5-ZMS1 INTERGENIC REGION.
GN YJR126C OR J2050.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA ROSE M., KOETTER P., EMTIAN K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: TO MAMMALIAN TRANSFERRIN RECEPTOR PROTEIN AND
CC PROSTATE-SPECIFIC MEMBRANE ANTIGEN (PSM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z49626; CA89657.1; .
DR KX Hypothetical protein.
SQ SEQUENCE 811 AA: 92017 MW: 4F94BBF1 CRC32;

Query Match 10.4%; Score 438.5; DB 1; Length 811;
Best Local Similarity 25.6%; Pred. No. 4.2e-24;

```

Matches	182; Conservative	109; Mismatches	254; Indels	167; Gaps	32;
Qy	24	ORVSPKRGHLEF-----EEDGEGEAETLAHCPMELRGP-----PLGSRP-	68		
Db	7	ERKREKEGGOLKERTVVMADPDNDNEAEVGTQLOQYSGEETTRDNEESMNDSEFTLLSRNNG	66		
Qy	68	QPNLPPMAAG-----RRAAYLVLTALLIFETGAFLGYVAFFRGCACGDSVLVY	118		
Db	67	RSNTISSTVSGEIMKEHMDKEKEMFYLLASLLYMG-FVAAF-APRIS-----L	114		
Qy	119	SEDVNYEPDLDPEHGRGLYMSDLOAMFLOPLGEGRLDTRIOTSLREYVAGSAG-MAALTQ	177		
Db	115	SRDFR-----RHSRSLTNAEYRIRYLNSLOQ-----ENRAKEHYKYGAGYMSNGAS	161		
Qy	178	DIRALSHROKLDHYMTDHY---VGLQFP---DPAHPNLTLMHWDEAGCEQLEPDDPV	231		
Db	162	D--SSTFYTTLDE-FLDNGYKPKYKERYIPGEIPDVTNAPL--ENGKYYTASMTEDRY	216		
Qy	232	-----YCPYSAIGNVTGELYAHYHGRPEDLODLRANGVDPVGRLLVRGVY	277		
Db	217	KGDPAASHARKROKGFHQYSKNGSVTARVYFCMYGSIIDYKLLKKNIDIEDKIHVRSK	276		
Qy	278	ISFAKVTNNADDFGAGVILYPEPD-----FSQDPK-PSLSSQAYYGAHV	324		
Db	277	ILPLGLKVNAAELGYASSVLIITYPDDEGKYTEENGFLAHPYGPARNPSTYIRDSV-NYFS	335		
Qy	335	LGTGDPYTPGPFPS-FNQCQFPVVASGCLPSIPAOPISADIASRLRLKLGVPARQEWOG	382		
Db	336	DHPGPTTPGYTSGSKSDPEHMSPVCR-VPRISPVMASRDQPIERLN-----	384		
Qy	383	SLGSPFYLGGPRL-----RLVYNNHRT-----STPNNIFGCLIEG	419		
Db	384	--GGFOIGPSPNKKDFGSFTGPSSSIDKVLHNLITINIKEMSSVEYSIGIF--TEG	438		
Qy	420	RSEPHYVYVIGQRPAMGPGAAKSA-VGTAILLLEYVTRPSSVWSNGEPRRSLTF-SMDG	478		
Db	439	E-----IITGHRSLASSSGAGDANSSGAILLETARGMSKILKRGWAKLPRIKLISMDG	492		
Qy	479	GDFGSGVSTEWLEGYLSYLAKAAVYVSLDNNAVLDGDKFHAFTSPLLISLIESVLKQYDS	538		
Db	493	ERSGLIGSTDYEAHAAILRRALRYLYLNDMAISGTN-PhcKANPLDQVITYEAAKLTGF	551		
Qy	539	PNHSQITLYEQVVFNPNSMDAIVIRPLPDSASYSTAVGVPAVEFSMEDQAYPLH	598		
Db	552	NGHEWDSLEFDHAKYTSNA---TISLLQGLSSTYSEFQYHLGVPAAHFOFNADTSGAYVH	607		
Qy	599	TKE-----DTYEN---LHKVLQGRLLPAAQAAVQALQGLLIRLSHDL	638		
Db	608	SNSVFDSPTWLEKFTNSDYKLLHNTM-----AMVGLTITLMSNEL	648		
RESULT 4					
YWAD_BACSU	YWAD_BACSU	STANDARD:	PRT:	455 AA.	
RC	P25152:				
DT	01-MAY-1992 (Rel. 22, Created)				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DT	01-NOV-1995 (Rel. 32, Last annotation update)				
DE	HYDROTHERMAL 49.5 KD PROTEIN IN DAE-TYRZ INTERGENIC REGION PRECURSOR.				
GN	YWAD OR IPA-8R.				
OS	Bacillus subtilis.				
CC	Bacteria: Firmicutes: Bacillus/Clostridium group;				
OC	Bacillus/staphylococcus group; Bacillus.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-168;				
RX	MEDLINE; 95020537.				
RA	GLASER P., KUNST F., ARNAUD M., COUDART M.P., GONZALES W.,				
RA	HULLO M.F., IONESCU M., LUBOCHINSKY B., MARCELINO L., MOSZER I.,				
RA	PRESECN E., SANTANA M., SCHNEIDER E., SCHWEITZER J., VERTES A.,				
RA	RAPOPORT G., DANCHIN A.;				
RT	"Bacillus subtilis genome project: cloning and sequencing of the 97				
RT	kb region from 335 degrees to 333 degrees."				

DB 364 SLTGGSSDHVPFHEAGIDSANFTW-----GDP-----ETEEVEPMWHPEDSIEHISK 412

RESULT 5
ID AMPX_VIBPR STANDARD; PRT; 504 AA.

AC 001693;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE BACTERIAL LEUCYL AMINOPEPTIDASE PRECURSOR (EC 3.4.11.10).
OS Vibrio proteolytica (Aeromonas proteolytica).
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 107-136 AND 233-405.
RC STRAIN=AFCC 15338;
RX MEDLINE: 92329552.
RA VAN HEKE G., DENSLON S., MATKINS J., WILSON K., WAGNER F.;
RT "Cloning and nucleotide sequence of the vibrio proteolyticus
RT aminopeptidase gene";
RL Biochim. Biophys. Acta 1131:337-340(1992).
RN [2]
RP SEQUENCE OF 73-504 FROM N.A.
RX MEDLINE: 92235063.
RA GUENET C., LEPAGE P., HARRIS B.A.;
RT "Isolation of the leucine aminopeptidase gene from Aeromonas
RT proteolytica. Evidence for an enzyme precursor";
RL J. Biol. Chem. 267:8390-8395(1992).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 107-396.
RX MEDLINE: 94373500.
RA CHEVRIER B., SCHALK C., D'ORCHYMONT H., RONDEAU J.M., MORAS D.,
RA TARNOUS C.;
RT "Crystal structure of Aeromonas proteolytica aminopeptidase: a
RT prototypical member of the co-catalytic zinc enzyme family.";
RL Structure 2:283-290(1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY COMPLEXED WITH INHIBITOR.
RX MEDLINE: 96215434.
RA CHEVRIER B., D'ORCHYMONT H., SCHALK C., TARNOUS C., MORAS D.;
RT "The structure of the Aeromonas proteolytica aminopeptidase complexed
RT with a hydroxamate inhibitor. Involvement in catalysis of Glu151 and
RT two zinc ions of the co-catalytic unit";
RL Eur. J. Biochem. 237:393-398(1996).
CC -1- CATALYTIC ACTIVITY: ACTS MOST RAPIDLY ON L-LEUCYL-PEPTIDES.
CC AMIDE AND BETA-NAPHTHYLAMIDE. DOES NOT CLEAVE GLU- AND ASP- BONDS.
CC -1- COFACTOR: BINDS TWO ZINC IONS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M28.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z11993; CAAT8039.1; -
DR EMBL: M85159; AAA21940.1; -
DR PIR: S24314; S24314.
DR PIR: S21684; S21684.
DR PDB: 1AMP; 31-AUG-94.
DR PDB: 1IGB; 01-AUG-96.
KW Hydrolyase; Aminopeptidase; Signal; Zymogen; Zinc; 3D-structure.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 106 POTENTIAL.
FT CHAIN 107 405 BACTERIAL LEUCYL AMINOPEPTIDASE.
FT PROPEP 406 504 REMOVED IN MATURE FORM (POTENTIAL).
FT METAL 211 211 ZINC (POTENTIAL).
FT METAL 214 214 ZINC (POTENTIAL).
FT METAL 336 336 ZINC (POTENTIAL).
FT DISULFID 329 333

FT ACT SITE 257 257 GENERAL BASE.
FT CONFLICT 303 304 TD -> DT (IN AA SEQUENCE).
FT CONFLICT 306 306 N -> D (IN AA SEQUENCE).
FT CONFLICT 312 314 TOL -> QT (IN AA SEQUENCE).
SQ SEQUENCE 504 AA; 54232 MW; 36911698 CRC32;

Query Match 3.4%; Score 143; DB 1; Length 504;
Best Local Similarity 23.1%; Pred. No. 0.0063;
Matches 72; Conservative 37; Mismatches 105; Indels 98; Gaps 16;

QY 344 PVY-----ASSGSPISPAQISADIAS-----RLKLKLGPAV-----PEMW-----GSL 385
DB 108 PPTIQQATVTAMLPQVDASOITGTSLSSEFTNFTTTSAGQASDIASEWALSLASL 167
QY 386 GSPYHLCGPRRLVYVNNHSTPNNIFCIEGRSEPDHYVYGAORDAW----- 437
DB 167 -----PNASVKGVSISGYNQ---KSVMTITGSEAPDEMIYIGHLDTIGSHTEQS 216
QY 437 -GPGAAKSAVGTAILLELVTFSSWNSNGFRPRSLLFISWDGDFGSVSTEMLEGYLS 495
DB 217 VAPGADDADAGIAVTEVIVLS---ENNFPKRSIAFMAYAAEEVGLRESODLANQY-- 272
QY 496 VLHKAIVVYSLDNAVLDGDKFPAKTSPLTSLSEVLRQVDSPNHSQTLREGVVFTNP 555
DB 272 -----KSEGRN-----VVSAL-GLDMTRYGSA--DDVVEITD 301
QY 556 SMDAEVIRPLP-----MDSAYSSTAF-----VGPAV--EFSEMEDDOAY 594
DB 302 YTDNFTFOYLTQMLDELPLSLITGFDICGYACSDHASMHNAGYPAAMPFSKFNQY--- 359
QY 595 PFLHTKEDTYEN 606
DB 359 PRHTTODTLAN 370

RESULT 6
SUBV_BACSU
ID SOBV_BACSU STANDARD; PRT; 806 AA.

AC P29141;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MINOR EXTRACELLULAR PROTEASE VPR PRECURSOR (EC 3.4.21.-).
GN VPR OR IPA-45R.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 161-195.
RX MEDLINE: 92041574.
RA SLOMA A., RUFO G.A. JR., THERIAULT K.A., DWYER M., WILSON S.W.,
RA PERO J.;
RT "Cloning and characterization of the gene for an additional
RT extracellular serine protease of Bacillus subtilis";
RL J. Bacteriol. 173:6889-6895(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE: 95020537.
RA GLASER P., KUNST F., ARNAUD M., COUDART M.P., GONZALES W.,
RA HULLO M.F., IONESCU M., LOBOCHINSKY B., MARCELINO L., MOSZER I.,
RA PRESECAN E., SANTANA M., SCHNEIDER E., SCHWEIZER J., VERTES A.,
RA RAPOPORT G., DANCHIN A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees";
RL Mol. Microbiol. 10:371-384(1993).
CC -1- FUNCTION: NOT REQUIRED FOR GROWTH OR SPOULATION.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- PTM: PROBABLY UNDERGOES C-TERMINAL PROCESSING OR PROTEOLYSIS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M76590; AAA22881.1; -
 DR EMBL: X73124; CA51601.1; -
 DR EMBL: 299123; CAB15835.1; -
 DR PIR: A41341; A41341.
 DR HSP: P00782; ISDA.
 DR SUBTILIST: Bg10591; VPR.
 DR PFAM: PF00082; Peptidase_S8; 2
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 DR Hydrolyase; Serine protease; zymogen; Signal.
 FT SIGNAL 1 28
 FT PROPEP 29 160
 FT CHAIN 161 806
 FT ACT_SITE 189 806
 FT ACT_SITE 233 806
 FT ACT_SITE 534 806
 FT ACT_SITE 534 806
 FT SEQUENCE 806 AA; 85608 MW; 25397F8F CRC32;

Query Match 3.1%; Score 130; DB 1; Length 806;
 Best Local Similarity 21.8%; Pred. No. 0.11;
 Matches 101; Conservative 52; Mismatches 163; Indels 148; Gaps 23;

QY 81 AAPYLVLTLTLFTGFLGFLGYAFRSCQAGSVLYVSEDEVY-EPDL--DHQGRLLW 137
 DB 164 SAPLYIA-----NDAMWDLGYT-----GKIKVAIIDTGEYVHPDKKNFGYRGY- 210
 QY 138 SDOAMFLQFL-----GGRLEDTIRGTSLSRERVASG----- 171
 DB 210 -----DFVNDYDPKKEPTGDPGEGATDHGTHAGVYAANGTICKVAPDAILAYRV 261
 QY 171 ---GMAALTQDIRAALSROKLDHWMTDTHYVG--IQPDPAPHNTLHWVDEAGKV----- 221
 DB 262 LGFGSGGTENYIAGVERAVQDADVWNLISGNSLNNPDWATSTALDWMASEGVAVTNSN 321
 QY 221 -----GEOLPLEDDPV-YCPYS-----AIGNVT 242
 DB 322 GNSGNGMTVSGFISREASISGATQLPLNEYAVTFGSYSASAKVMYKNEEDVKALNKE 381
 QY 243 GELVYAHYGRPEDLDLARGVDPVGRLLLVYGVISFAQKVTNAODFGAGVLIYF-P 301
 DB 382 VELVEAGIEAKQFE-----GKDLTGKVAVYKKGSIATFVDKADNAKAGALGMVYVNNIS 436
 QY 302 ADEFSQDP-----KPSLSSQQAIVGHVHLGTGDPYTPGFSPFNQTFPPVASSGLPSI 354
 DB 437 GELIENVVPMSPYPTIKLSLEDGEKLVSAKAGE-----TKTFEKLIVSKAL--- 483
 QY 355 PAQPIASDASRLRLKRLKLPVAPQEQSGSLGSPYHLGGRPLRLVYNNHRTSTPINNIF 414
 DB 483 -GEQV-ADFSR-----GVM-DTW---MKPDISAPGVNVTSTPTHPDPHYG--Y 527
 QY 415 GCIEGRSEPDHYV-----VIGAQRDAMGPGAAKSAV-GTALLLE 452
 DB 528 GSKQGTSMASPHIAGAVAVIKAKPKMSVQIKAAIMNTAVTLK 571

RESULT 7
 ID V12_BPV2 STANDARD: PRT: 467 AA.
 AC P06457;
 DT 01-JAN-1988 (rel. 06, Created)
 DT 01-JUL-1989 (rel. 11, Last sequence update)
 DT 01-OCT-1996 (rel. 34, Last annotation update)
 DE MINOR CAPSID PROTEIN L2.

GN 12.
 OS Bovine papillomavirus type 2.
 OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA GROFF D.E., MITRA R., LANCASTER W.D.;
 RL Submitted (MAY-1988) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 85107102.
 RX POTTER H.L., MEINKE W.J.;
 RT "Nucleotide sequence of bovine papillomavirus type 2 late region";
 RL J. Gen. Virol. 66:187-193(1985).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M20219; AAA66839.1; -
 DR EMBL: X01768; CAA25907.1; -
 DR EMBL: M24326; AAA46924.1; -
 DR PIR: B31169; P2WIB2.
 DR PFAM: PF00513; Late-protein_L2; 1.
 DR Coat protein; Late protein.
 FT CONFLICT 281 281
 FT CONFLICT 306 306
 FT CONFLICT 349 349
 FT CONFLICT 377 377
 FT SEQUENCE 467 AA; 49523 MW; 22E71F79 CRC32;

Query Match 2.9%; Score 123.5; DB 1; Length 467;
 Best Local Similarity 24.4%; Pred. No. 0.15;
 Matches 70; Conservative 35; Mismatches 99; Indels 83; Gaps 16;

QY 120 EDVNEPDLDFHQRLLYWSLDQAMFLQFLGEGRELDTRIOTSLRERV-AGSAGMALRTOD 178
 DB 179 DHANQVSNAAVHGSAYHAPLD-----LQSSIAETSGLEINIFVGGAGLG----- 223
 QY 179 IRAALSROKLDHWMTDTHYVGIGQFPDPAPHNT-----LHWVDEAGKVBQLP 225
 DB 223 -----DTGGENIELTFPGSPRTSTPRNLQFARGILNMFSK--RYTQIP 265
 QY 226 LEDPDV-----YCPYSAI-----GNVYGEVYAHYGRPEDLDLARGVDPVGRLL 270
 DB 266 TEDPDVFSQTSNPNLYDEPAVLKGPSSGVRGLSGVY---RPDIE---TRGGGVGPQ 318
 QY 271 LLVRVGVISFAQKVTN---AODFGAGVLIYF---EPADFQDPKPPSLSSQQAIVGHVH 324
 DB 319 LHVRYSLSTTTEDEVAIPAVDEDETOGLAFLPILHEPQDF-DEILDLDGGERHALLPKSY 377
 QY 325 ---LGTG--DPTTPGFSPFNQTFPPVASSGLPSI-PAQPIASDAS 365
 DB 378 TAPISGVARRALIPG-QGFSATRPFGVVTYGGSDPMVPASVGPDPST 423

RESULT 8
 ID TRC5_ECOLI STANDARD: PRT: 1448 AA.
 AC P27190; P27191; P27185;
 DT 01-AUG-1992 (rel. 23, Created)
 DT 01-AUG-1992 (rel. 23, Last sequence update)
 DT 15-DEC-1998 (rel. 37, Last annotation update)
 DE DNA PRIMASE TRAC (EC 2.7.7.-) (REPLICATION PRIMASE).
 GN TRAC.
 OS Escherichia coli.
 OG Plasmid Incp-beta 751.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.

FT DOMAIN 197 314 COIL 2A.
SQ SEQUENCE 1805 AA; 198744 MW; 789C2B48 CRC32;

Query Match 2.7%; Score 112.5; DB 1; Length 1805;
Best Local Similarity 22.0%; Pred. No. 7.2;
Matches 106; Conservative 44; Mismatches 149; Indels 183; Gaps 27;

QY 16 PSSQTVYRVGPKR--GHLEEEDEDEG-----AETLAHFCMELRGPPLGSR 67
DB 1447 PAGESALREVENEPFEGEIPGEGQDMEEGESEADLGETLP---DSTPLDLYR 1502
QY 68 QNLFPEMAAG--RRAPYLVLTALIFGAFLLGYAFRGSCQAGDS-----VLVYS 119
DB 1503 SPASPRWDLAGEORLSPQ-----GDAKEDMGPAVPPA 1535
QY 120 EDVNEPDLDFHQRGLYMSDLQAMFLQFLGEGRLDTIRQTSLRERVAGSAGMALVTDI 179
DB 1536 QGLSGPPEEEEG--HGSDLSS-----EEFDLGEASLLPGV----- 1573
QY 180 RAALSRQKLDHYMTDTHYVGLQEPDPAHNTLHW-----VDEAGKVEQLPLED 228
DB 1573 ---PDEVADHVG-----GVPPVLAQAC--WDQGESDGFADDEEGEGEE---ED 1615
QY 229 PUYGPRYSAGNVTGELYAH--YGRPEDLDLRARGVDPVGRLLVRGVISFAQKYN 286
DB 1616 AD-----EAGASGQMWMSGASGGCKVQDIAQRG-DPVQE---SVG-----VSG 1657
QY 287 ADFGAQGVLT-YPEPADFSQDPKPKFSL-----SQAAVYGHV--HL-----G 326
DB 1658 LWDGDLRGAAAVPALEMYSDSAEPGSESEASLSEEGQVTDHLDAPQEVTSWVG 1717
QY 327 TGDPTPG--PFSFNQTPPPVASSGLSPISIPQPI-SADIASLLRK-----LKPVP 377
DB 1718 VDAEDDIGQSPNLDSEQVNGKMEGLEQAEQVLDGDELDLQGEVGAALKVP--- 1775
QY 378 QEMOGLSGPYHLGPGPRLRLVNNHRTSTPINIFGCIEGSEPDHVVIGAQDANG 437
DB 1775 ---LVASPVHLGP-----SPLKRTLLSGVDGDSMS 1801
QY 438 PG 439
DB 1802 SG 1803

RESULT 10
P2P_LACPA STANDARD: PRT; 1902 AA.
AC Q02470;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-
ASSOCIATED SERINE PROTEINASE) (Lp151).
GN PRTP.
OS Lactobacillus paracasei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCDO 151;
RX MEDLINE; 92381481.
RA HOLCK A., NAES H.;
RT "Cloning, sequencing and expression of the gene encoding the cell-
envelope-associated proteinase from Lactobacillus paracasei subsp.
paracasei NCDO 151".
RL J. Gen. Microbiol. 138:1353-1364(1992).
RN [2]
RP SEQUENCE OF 189-196.
RX MEDLINE; 92226694.
RA NAES H., NISSEN-MEYER J.;
RT "Purification and N-terminal amino acid sequence determination of the
cell-wall-bound proteinase from Lactobacillus paracasei subsp.

RT paracasei";
RL J. Gen. Microbiol. 138:313-318(1992).
CC -1- FUNCTION: PROTEASE WHICH BREAK DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC INSULIN B-CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M83946; AAA25248.1; -;
DR PIR; B44858; B44858.
DR HSSP; 099405; IMPT.
DR PFAM; PF00082; Peptidase_S8; 3.
DR PFM; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_SER; 1.
DR PROSITE; PS00138; SUBTILASE_HIS; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE NEG.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal;
KW Transmembrane.
KM SIGNAL 1 33
FT PROPEP 34 187 POTENTIAL.
FT CHAIN 188 1902 PII-TYPE PROTEINASE.
FT DOMAIN 188 1876 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1877 1895 MEMBRANE ANCHOR (BY SIMILARITY).
FT DOMAIN 1896 1902 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1902 AA; 200253 MW; 932EC634 CRC32;

Query Match 2.7%; Score 111.5; DB 1; Length 1902;
Best Local Similarity 21.3%; Pred. No. 9.2;
Matches 100; Conservative 63; Mismatches 164; Indels 143; Gaps 23;

QY 214 VDEAGKVEQL-----PLEDDVYCPYSAIGNVTGELYA----- 249
DB 335 IEDSAKIGADVILNMSLGSQNGQLEDEDEIAVQANASGTAAVASGNSGTSGSATGV 394
QY 249 ---HYGREPDQDLARAVDPVGRLLLRVGVISPAQVNTAAQ--DFAQGVLYLPEPAD 303
DB 395 NKDYTG---LQDNEMVGT-----GTSRCATTVASAENMTDVLSQAVTI--TDK 439
QY 304 FSDPPKPSLSQQAUVYGHVHLGTDPYTPGPFNSNQTQPPV--ASSGLSPISPAQISA 361
DB 440 LQLGPEITQLSNDP-----TG-----SPDQKFFVVKDASGDLKGAADTA 483
QY 362 DIASRLKLKLGPAVAPQEMQSLGSPYHLGPGPRLRLVNNHRTSTPINI----- 414
DB 484 DAKGIAIVKRGELNFAQKQ-----KYAQAAGAAGLIIVNNDGATGATLSIRLTTPPT 537
QY 414 FGCIGRSEPDHYVYVIGAQRAMGGAASAVGATLILELVRT-----ESSMYSNGFR 466
DB 538 FG-LSSKT-----GQKLVDMWTAHPDDSLGVKIALTLIPNOKYTEDKMSDFSTYG-- 587
QY 467 PRSLTF---LSWDGDDGSGVSTEWLEGYLS-----VLHKAVVYVSDNAVIG 513

DB 587 PVSNTSEKPDITAPG---GNINSTONNGYTNKSGTSMASPIAGSOALLKQALNK--- 641
 QY 514 DDKFAKTSPLLSLIESVLKQVD-----SNHGQILYEVVTFNPS 556
 DB 641 NNEFYADYQOLKGTALTDLTKTEKMTAOPINDINYNNAVSVRRQAGI---VVKAA 696
 QY 557 WDAEVRPLPMDSSAFTAFVPAVEF-SFEMDDOAVFLTKDQTYE 605
 DB 697 IDA-----LEKNPSTVVAENGYPAVELKDTSTDKTKLTFTNTTHE 739

RESULT 11
 SCPA_STRPY
 ID SCPA_STRPY STANDARD: PRT: 1167 AA.
 AC P15926;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE C5A PEPTIDASE PRECURSOR (EC 3.4.21.-) (SCP).
 GN SCPA.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 CC Streptococcus.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-41.
 RX MEDLINE: 90153964.
 RA CHEN C.C., CLEARY P.P.;
 RT "Complete nucleotide sequence of the streptococcal C5a peptidase gene
 of Streptococcus pyogenes."
 RL J. Biol. Chem. 265:3161-3167(1990).
 CC -1- FUNCTION: THIS VIRULENCE FACTOR OF S. PYOGENES SPECIFICALLY CLEAVES
 C-TERMINUS, DESTROYING ITS ABILITY TO SERVE AS A CHEMOTACTICANT.
 CC -1- SUBCELLULAR LOCATION: CELL WALL BOUND.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 SUBTILASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: J05229; AAA26960.1; -
 DR PIR: A35066; A35066.
 DR HSSP: P00782; 2SBR.
 DR PFAM: PF00082; Peptidase_S8; 4.
 DR PFAM: PF00746; Gram_pos_anchor; 1.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
 KW Hydrolase; Serine protease; signal; Cell wall; Repeat; Transmembrane.
 FT SIGNAL 1 31
 FT CHAIN 32 1167 C5A PEPTIDASE.
 FT DOMAIN 32 1139 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1140 1157 POTENTIAL.
 FT DOMAIN 1158 1167 CYTOPLASMIC (POTENTIAL).
 FT ACT_SITE 130 130 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 193 193 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 512 512 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DOMAIN 1029 1104 CELL WALL ATTACHMENT DOMAIN (POTENTIAL).
 FT DOMAIN 1034 1050 1. 4 X 17 AA TANDEM REPEATS.
 FT REPEAT 1051 1067 2.
 FT REPEAT 1068 1084 3.
 FT REPEAT 1085 1101 4.
 FT DOMAIN 1127 1131 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.
 FT SEQUENCE 1167 AA: 128263 MW: 5816687B CRC32:

Query Match 2 6%: Score 108.5; DB 1; Length 1167;
 Best Local Similarity 18.8%: Pred. No. 7.3;
 Matches 139; Conservative 90; Mismatches 227; Indels 285; Gaps 37;

QY 30 RKGHLEEEDEDEGEAGETLAHFCPMLRGPEPL-GSRPQPNLIPMAAGRAAPLYLT 88
 DB 183 KQKTAIVDQEHGTHSGILSGNAPSETKEPYRLEGAMPE-----AQLLLM 227
 QY 89 ALLIFGAFILGVAFRGSCQAGDSVLVSDVNYEPDLDFHOGRLYWSDLQAMFLQFL 148
 DB 228 RVEIYNG--LADYA--RNAQAIRDAVNLGAKYIN---MSGNALAVANLP----- 273
 QY 149 GEGRELEDTIRQTSLRERVAGSAGMAALTODITAAISRKQLDHWIDTTHYG--LQFPDPAH 207
 DB 273 -----DETKKAFDYAKSKGVSIVTS-----AGNDSFGGKTRLPALAH 310
 QY 208 PNTLHWVDAGVYG-----PADP-----SQDPKPSLSSQOAYGVHVLGTGDPY 331
 DB 311 P-----DYGVGIPPAADSTLTVAISYSPDKQLETANVKITDDQDKEMPVLSTRFEP 363
 QY 235 YSAIGNVTGELYAHYGRPELDLRLARGVDVGRLLVRVGVISPAQKVTNAODFGAOG 294
 DB 364 ----NKAYDYAVANGMKED--DFK----DYKGIALLERDIDIDKDKVANAKKAGAVG 412
 QY 295 VLITYE-----PADP-----SQDPKPSLSSQOAYGVHVLGTGDPY 331
 DB 413 VLYIVNQDKGFPIELPNVDMQPAAFISRKDGLLKXNPQKITYFNA----- 459
 QY 332 TEG-FPSFNQTFPPVYASGL-----PSI--PAQPIADIASRLRLKGPVAPQEWQ 381
 DB 459 TPRLVPLTASGTRLSRFSWGLTADGNIKPDIAAGODILSVANNKYALSGTSMAPLV 518
 QY 382 GSLIG-----SPYHLGPG----- 395
 DB 519 AGIMGLQKQYEQYPDMPSERLDLAKKVLMSATALLVDEDEKAFSPRQGAGAVDAK 578
 QY 395 ---PRRLVYNNHRTSTPL--NNIFG-----CIEGRSE--PDHYVITAGRD----- 435
 DB 579 KASATMYVTDKDMTSSKVLNVSDFEVTYVHNKSKDPOLYYQATVQTDKVDGKHF 638
 QY 435 AMQPGA-----AKSAVGTALLLEIVRTFS--SNWSNPF-----RP 467
 DB 639 ALAPKVLVYASWQKITIPANSSKQVTPIDASR-FSKDLAQKKNQYFLEGVYRFKQDPT 697
 QY 468 RRSLLFISWDG--GDFGSYSTEWLEGYLVHLKAVVYVSIDNAVLGDKDFHAKTSPLL 525
 DB 698 KEELMSIPYIGFRGDFGNLSAVE-----KPIYDSKD-----GSSYHEANSDAK 741
 QY 526 TSLIESVLKQVDSPNHSGOTLYEQVVFVNPSWDAEYIRPL-----PMDSSAFTAF 577
 DB 742 DQIDGDGLOFYALKNNFTALTTE---SNP-W--TIKAVKEGVENIEDIESESEITETIF 794
 QY 578 VGVPAVEFSFEMDDOAVYPLH 598
 DB 795 AGTFA-----KODDSHYITH 810

RESULT 12
 EFG2_TREPA
 ID EFG2_TREPA STANDARD: PRT: 695 AA.
 AC 083748;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE ELONGATION FACTOR G 2 (EF-G 2).
 GN FUSB OR FUSA-2 OR TP0767.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NICHOLS.
 RX MEDLINE: 98332770.

FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 887 TYROSINE-PROTEIN KINASE RECEPTOR UFO.
FT DOMAIN 19 442 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 443 465 POTENTIAL.
FT DOMAIN 466 887 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 29 118 IG-LIKE DOMAIN.
FT DOMAIN 134 205 IG-LIKE DOMAIN.
FT DOMAIN 217 315 FIBRONECTIN TYPE-III.
FT DOMAIN 319 416 FIBRONECTIN TYPE-III.
FT DOMAIN 529 800 PROTEIN KINASE.
FT NP_BIND 535 543 ATP (BY SIMILARITY).
FT BINDING 560 560 ATP (BY SIMILARITY).
FT ACT_SITE 665 665 BY SIMILARITY.
FT MOD_RES 696 696 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT VARSPIC 422 430 MISSING (IN SHORT FORM).
FT CONFLICT 296 296 P -> T (IN REF. 2).
FT CONFLICT 331 331 E -> K (IN REF. 2).
FT CONFLICT 632 632 G -> D (IN REF. 2).
SQ SEQUENCE 887 AA; 97374 MW; E747BA73 CRC32;

Query Match 2.5%; Score 107; DB 1; Length 887;
Best Local Similarity 21.1%; Pred. No. 6.2;
Matches 118; Conservative 66; Mismatches 142; Indels 232; Gaps 33;

QY 282 QKVTNAQDQAGVLIYPPADFSODPPKPSLSQQAAYGH-VHGTGDPYTPGF----- 336
DB 254 QAVLSDQDNGIOA---GSP-----DPPEPLTSQASVPPHOLRLGSLPHPPYHIRVAC 304
QY 336 -----PSFNQTOFPYVYAS-SGLPSIPAOPIASD-INSRLRLKGVYAOQMOGSLGSP 388
DB 305 TSSQGPS-SWTWMLPEVTEEGVPLGPPENISATRNGSQAFVHQEPRAP-LOGTILG-- 360
QY 389 YHGG---GPRRLRVVNNHRTST-----PINNIFGCIENRSEPDHYVIG----- 431
DB 360 YRLAOGQOTPEVMDIGLROEVTLELQDGSVSNLTVCVAA-----YTAAGDGWSLP 413
QY 431 AORDAMGPPAAK-----SANGTALILELV----- 455
DB 414 VPLEAMRPEADPVHQLVKEPSTPAFSWPMWYVLLGAAVVAACVLLIALFLVHRRKKEFR 473
QY 455 -RTFSWMSNG-----FPRRSL-----LFISM----- 478
DB 474 YGEVEPTVERGELVYRVKRSYSKRTTEATINSIGISELKEKRDVMDRHKVALCK 533
QY 478 ---GGDFGSGVSTEWLEGYLVHLKAVYVSLDNNALVLDDEKPHAKT-----SPLLT 526
DB 534 TLGEGFGAV-----MEGL-----NODDSIL---KVAVTKMIAICTRSELED 574
QY 527 SLIESV-LKQVSPNHSQGLVEQVYFTNPMDAIVIRPLM---DSAYSTAFVGVFA 582
DB 575 FLSEAVCMKEFDHPN-----VMRLIGVCFQOSESESPFA---FV 610
QY 583 VEESFEMDQAVPELHTEKEDYENLHKVLOGRLPAVAQAAGOLLRLSHDRLLPD 642
DB 611 VILPEFKHDDLHSL-----LTSRLGGQ-----PVILPTMLVKNADIASGWE 654
QY 643 FGR---YGDVYLRIHIGNLNE-----FSGDLKARG---LTLQWYSAR 678
DB 655 YLSTRKFIHRLAARNC-MLNEMSMCVADFGLSKRIYNGDYRQRIKMKPKWI----- 710
QY 679 GDYIRAEKLRQETISSE 696
DB 710 -----AIESIADRVTYSK 722

RESULT 14
PIP_LACIC STANDARD; PRT; 1902 AA.
AC P16271;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE PI-TYPE PROTEINASE PRECURSOR (EC 3.4.21.-) (WALL-ASSOCIATED SERINE
DE PROTEINASE).
GN PRP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pW05.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
RM [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WG2;
RX MEDLINE; 88149035.
RA KOK J., LEENHOUTS K.J., HAANDRIKMAN A.J., LEDEBOER A.M., VENEMA G.;
"Nucleotide sequence of the cell wall proteinase gene of
Streptococcus cremoris Wg2".
RL Appl. Environ. Microbiol. 54:231-238(1988).
CC -1- FUNCTION: PROTEASE WHICH BREAK DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC INSULIN B-CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: BELONGS TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24767; AAA17677.1; -.
DR HSRP; Q99405; IMPT.
DR PFM; PF00082; Peptidase_S8; 3.
DR PFM; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KM Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KM Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1902 AA; 199910 MW; FDE60D93 CRC32;

Query Match 2.5%; Score 106.5; DB 1; Length 1902;
Best Local Similarity 20.3%; Pred. No. 21;
Matches 96; Conservative 65; Mismatches 163; Indels 149; Gaps 22;

QY 214 VDEAGKVGSQL-----PLEDDPYVCPYAIGNVTGELVYA----- 249
DB 335 IEDSAKIGADVLMKSGDSNGOTLEDPELAAVQANESGTAAVISAGSGTSATEGV 394
QY 249 ---HGKRPEDLDDLRARGVDYVGR-----LLIVRGVLSFAQVTNADFGAGVLIYPE 300
DB 395 NRDDYG---LQDNEVNGPRTSGRATVVAASENTDVITQAVTIND---GTGLQIGP- 445
QY 301 PADFSODPPKPSLSQQAAYGHVHGTGDPYTPGFPSFNQTOFPYV--ASSGLPSIPAP 358

```

Db 445 -----GTQLSSND-FT---GSFDQKFFVVKDASGNLSKALAD 480
QY 359 ISADIASRLRLKKGPAVPOEMOGSLGSPYHLGPGRLVNNHRTSPINNI----- 414
Db 481 YTAADKAGKIAIVKRGELTFADKQ-----KYQAAGAAAGLIIVNDDGTATPVSALMTT 534
QY 414 ---FGCIEGRSEPDHYVYVIGAQRDAMGPAKSAVGAITLLELVRT-----FSSVSN 463
Db 535 FPTFGSS-----VTGOKLYDWAHAHPDSDLGVKIALTLVPNOKYTEDKMSDFTSY 585
QY 464 GFRPRRSILF---ISWDGDFGSGVSTEMLEGYS-----VLHKAIVVYSLDNA 510
Db 586 G--PVSLSFKRPDITAG---GNLWSTQNNNGYTNMGSTMASPFIAGSALLKQALNKK 640
QY 511 VLGDGKFAKTSPLTSLISLVKQVD-----SPNHSQOTLYEQVYFT 553
Db 641 ---NNPFYAYYKOLKGALTDLFKLTVEKMTAOPINDINYNVIVSPRQAGL-----VDV 693
QY 554 NPSMDAEVIRPLPMDSSAYSFATFVGPVAFEF-SFMEDDQAYFHLTKEDTYE 605
Db 694 KAALDA-----LEKNPSTVAENGIPAVELKDFSTDKFTKLTFTNRTTHE 739

RESULT 15
P2P_LACLA STANDARD: PRT: 1902 AA.
ID P2P_LACLA
AC p15293:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE P1-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCYPIN) (CELL WALL-
DE ASSOCIATED SERINE PROTEINASE) (LP151).
GN PRT.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OG Plasmid pLP763.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCDO 763.
RX MEDLINE; 89313288.
RA KIMURA M., IKEMURA H., SHIMIZU-KADOTA M., HIRASHIMA A.;
RT "Molecular characterization of a cell wall-associated proteinase gene
RT from Streptococcus lactis NCDO763."
RL Mol. Microbiol. 3:359-369(1989).
CC -1- FUNCTION: PROTEASE WHICH BREAK DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC INSULIN B-CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X14130; CAA32350.1; -.
CC PIR; S06997; S06997.
CC HSSP; P00782; 2S8T.
CC PRAM; PR00082; Peptidase_S8; 3.
CC DR PROSITE; PS00136; SUBTILASE_ASPI; 1.
CC DR PROSITE; PS00137; SUBTILASE_HIS; 1.
CC DR PROSITE; PS00138; SUBTILASE_SER; 1.

```

```

DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
FT SEQUENCE 1902 AA; 200139 MW; 8617ECFB CRC32;
PROTEINS.
QY 214 VDEAGKVEQL-----PLEDDPVYCPYSAGNVTGELYVA----- 249
Db 335 IEDSAKIGADVLMNLSGSDGNGQTFLEDPILAAYONANESGTAIVISAGNGSGSATEGV 394
QY 249 ---HYGRPELDLDRANGVPVGR-----LLIYRGVISTAQKVTANQDFGAGVLIYPE 300
Db 395 NKDYGG---LQDNEMVGTGTSGRATTVASAEITDITGAVTID---GTGLQGPPE 445
QY 301 PADFSQDPKPKPSSQOAVYGHVLTGDPYTPGPFPSFNQTOPPPV--ASSGLPSIPAQP 358
Db 446 TIQLSSND-----FTG-----SFDQKFFVVKDASGNLSKGVAD 480
QY 359 ISADIASRLRLKKGPAVPOEMOGSLGSPYHLGPGRLVNNHRTSPINNI----- 414
Db 481 YTAADKAGKIAIVKRGELTFADKQ-----KYQAAGAAAGLIIVNDDGTATPVSALMTT 534
QY 414 ---FGCIEGRSEPDHYVYVIGAQRDAMGPAKSAVGAITLLELVRT-----FSSVSN 463
Db 535 FPTFGSS-----VTGOKLYDWAHAHPDSDLGVKIALTLVPNOKYTEDKMSDFTSY 585
QY 464 GFRPRRSILF---ISWDGDFGSGVSTEMLEGYS-----VLHKAIVVYSLDNA 510
Db 586 G--PVSLSFKRPDITAG---GNLWSTQNNNGYTNMGSTMASPFIAGSALLKQALNKK 640
QY 511 VLGDGKFAKTSPLTSLISLVKQVD-----SPNHSQOTLYEQVYFT 553
Db 641 ---NNPFYAYYKOLKGALTDLFKLTVEKMTAOPINDINYNVIVSPRQAGL-----VDV 693
QY 554 NPSMDAEVIRPLPMDSSAYSFATFVGPVAFEF-SFMEDDQAYFHLTKEDTYE 605
Db 694 KAALDA-----LEKNPSTVAENGIPAVELKDFSTDKFTKLTFTNRTTHE 739

```

Search completed: January 13, 2000, 01:09:18
Job time: 135 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2000, 12:10:22 : Search time 39.36 Seconds

(without alignments)
1246.581 Million cell updates/sec

Title: US-09-358-755-1

Sequence: 1 MERLWGLFORAQLSPRSSQ.....TLGGANALSGDWMINDNF 801

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database: SPTREMBL_11:*

Word size: 0

Number of hits that pass the threshold: 199794

1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4034.5	96.0	780	4	075422	075422 homo sapien
2	1504.5	35.8	757	11	007891	007891 cricetus
3	1450.5	34.5	763	11	062351	062351 mus musculu
4	1424.5	33.9	622	11	099376	099376 rattus norv
5	1349.5	32.1	778	13	090997	090997 gallus gall
6	775	18.4	752	11	P70627	P70627 rattus norv
7	756	18.0	745	11	O54697	O54697 rattus norv
8	727	17.3	750	4	043748	043748 homo sapien
9	722	17.2	752	11	035409	035409 mus musculu
10	709.5	16.9	751	6	077564	077564 sus scrofa
11	621	14.8	751	5	P91406	P91406 caenorhabd
12	541	12.9	794	3	043023	043023 schizosacch
13	510	12.1	1483	5	093332	093332 caenorhabd
14	297.5	7.1	783	3	008919	008919 saccharomyc
15	271.5	6.5	277	11	061560	061560 mus musculu
16	243	5.8	809	3	008693	008693 saccharomyc
17	234.5	5.6	322	4	043176	043176 homo sapien
18	208	4.9	63	11	064187	064187 cavia porce
19	159	3.8	501	2	P96152	P96152 vibrio chol
20	155.5	3.7	493	5	076552	076552 acanthochei
21	154.5	3.7	485	2	053737	053737 streptomyce
22	134.5	3.2	472	11	0921Y1	0921Y1 rattus norv
23	132	3.1	393	2	082996	082996 aeromonas c
24	131.5	3.1	565	1	057813	057813 pyrococcus
25	128.5	3.1	835	2	069873	069873 streptomyce

26	126	3.0	566	5	093318	093318 caenorhabd
27	120.5	2.9	500	2	P96264	P96264 mycobacteri
28	117.5	2.8	374	3	004033	004033 saccharomyc
29	117.5	2.8	561	3	074125	074125 coprinus ci
30	116.5	2.8	1301	2	006307	006307 allicyclobac
31	116.5	2.8	1083	2	086637	086637 streptomyce
32	116	2.8	1017	2	005586	005586 mycobacteri
33	115.5	2.7	640	3	002495	002495 schizophy11
34	113.5	2.7	634	4	075114	075114 homo sapien
35	112.5	2.7	701	12	083131	083131 avian myelo
36	112.5	2.7	701	12	089810	089810 avian myelo
37	112	2.7	1128	1	051999	051999 halobacteri
38	112	2.7	433	11	070216	070216 mus musculu
39	109.5	2.6	2731	2	0923T9	0923T9 pseudomonas
40	108.5	2.6	1557	5	096652	096652 drosophila
41	108.5	2.6	1075	11	P97305	P97305 mus musculu
42	108	2.6	1150	2	053637	053637 streptococ
43	107.5	2.6	809	5	044391	044391 strongyloce
44	107	2.5	679	10	023352	023352 arabidopsis
45	106.5	2.5	2314	2	069822	069822 streptomyce

ALIGNMENTS

RESULT 1
ID 075422 PRELIMINARY; PRT; 780 AA.
AC 075422:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE TRANSFERRIN-RECEPTOR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA GLOECKNER G., ROSENTHAL A., SCHERER S., WEBER J., SCHATTEVOY R.,
RA TSUI L.-C.;
RT "Large scale analysis of two regions in human chromosome 7q22:
RT annotation of 650 kb of genomic sequence around the PCOLCE and CUTL1
RT loci reveals 17 genes."
RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF053356; AAC78796.1; -
SQ SEQUENCE 780 AA; 86621 MW; 84BC2BD2 CRC32;

Query Match 96.0%; Score 4034.5; DB 4; Length 780.
Best Local Similarity 95.9%; Pred. No. 0;
Matches 774; Conservative 0; Mismatches 0; Indels 33; Gaps 2;
QY 1 MERLWGLFORAQLSPRSSQTVYORVEGPKHLEEEDEGEAEFLAHFCPELRGPE 60
Db 1 MERLWGLFORAQLSPRSSQTVYORVEGPKHLEEEDEGEAEFLAHFCPELRGPE 60
QY 61 PLGSRPPOPLIWWAAGRAAPYLVTALLITGAFLLGYVAFRSCQACGSVLVSE 120
Db 61 PLGSRPPOPLIWWAAGRAAPYLVTALLITGAFLLGYVAFRSCQACGSVLVSE 120
QY 121 DVNYEPDLDFHGRRLYSDIQAMFLOFLGEGRLIEDTIRQTSLEPRVAGSAGMAALTODIR 180
Db 121 DVNYEPDLDFHGRRLYSDIQAMFLOFLGEGRLIEDTIRQTSLEPRVAGSAGMAALTODIR 180
QY 181 AALSROKLDHVTHTHYVGLQFPDPAPNPNTLHWVDAAGKVEQLPLEDDPVYPSAIGN 240
Db 181 AALSROKLDHVTHTHYVGLQFPDPAPNPNTLHWVDAAGKVEQLPLEDDPVYPSAIGN 240
QY 241 VTGELYVAHYGREPDLQDARGVDPVGRLLLVAVGVISFAQKVTNAODFGAGVLYIPE 300
Db 241 VTGELYVAHYGREPDLQDARGVDPVGRLLLVAVGVISFAQKVTNAODFGAGVLYIPE 300
QY 301 PADFSODPKPKSLSSQQAAYGVHVLCTGDPYTGPFPSFNOTOPPVASSGLPSIPAPDIS 360
Db 301 PADFSODPKPKSLSSQQAAYGVHVLCTGDPYTGPFPSFNOTOPPVASSGLPSIPAPDIS 360

```

Db 301 PADFSQDPKPSLSSQAVYGVHVLGTGDPYTPGSPFSNQD----- 343
OY 361 ADIASRLRLKLVPAPOEMOGSLGSPYHLGPGPRLRLVNNHRTSPINNIFGCIETR 420
Db 343 -----KLKGPAPQEMOGSLGSPYHLGPGPRLRLVNNHRTSPINNIFGCIETR 393
OY 421 SEPDIYVIGARQAMGGAASAVGTALILELVTFSSMNSNGRPRLSLFISWDGD 480
Db 394 SEPDIYVIGARQAMGGAASAVGTALILELVTFSSMNSNGRPRLSLFISWDGD 453
OY 481 FGSVSTEMLEGYLSVLHLKAVYVSLDNAVGLGDKFNAKTSPLTSLIESVLKQVDSN 540
Db 454 FGSVSTEMLEGYLSVLHLKAVYVSLDNAVGLGDKFNAKTSPLTSLIESVLKQVDSN 513
OY 541 HSGQTLVEOVFTNPSMDAEVIRPLPMDSSAYSTAFVGVPAVEFSFEMDDQAYPFLTK 600
Db 514 HSGQTLVEOVFTNPSMDAEVIRPLPMDSSAYSTAFVGVPAVEFSFEMDDQAYPFLTK 573
OY 601 EDTYENLHKVLOGRLPAVAQAVAGLQGLLRSLSDRLPLDFGRYGDVLRHIGNLNEF 660
Db 574 EDTYENLHKVLOGRLPAVAQAVAGLQGLLRSLSDRLPLDFGRYGDVLRHIGNLNEF 633
OY 661 SDDLKARGITLQWYSANGDYIRAEKLRQEIYSESEDERLTRMYNVRIMF-----VE 714
Db 634 SDDLKARGITLQWYSANGDYIRAEKLRQEIYSESEDERLTRMYNVRIMFISLAQVE 693
OY 715 FYFLSQYVSPADSPRHRIFMGGRDHTLGLDLHLRLNSNGSGTGATSSIGFQSRERR 774
Db 694 FYFLSQYVSPADSPRHRIFMGGRDHTLGLDLHLRLNSNGSGTGATSSIGFQSRERR 753
OY 775 QLALLTWLQGANALSGDVNMDNNEF 801
Db 754 QLALLTWLQGANALSGDVNMDNNEF 780

RESULT 2
007891 PRELIMINARY: PRT: 757 AA.
AC 007891:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE TRANSFERRIN RECEPTOR PROTEIN (TR)
OS Cricetus griseus (Chinese hamster)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetus.
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE-Ovary:
RX MEDLINE: 94012749.
RA COLLAUN J.F., LAI A., DOMINGO D., FITCH M., HATTON S.,
RA TROMBRIDGE I.S.;
RA "YTRF is the conserved internalization signal of the transferrin
RT receptor, and a second YTRF signal at position 31-34 enhances
RT endocytosis";
RL J. Biol. Chem. 268:21686-21692(1993).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE: 90226333.
RA ALVAREZ E., GIRONES N., DAVIS R.J.;
RA "A point mutation in the cytoplasmic domain of the transferrin
RT receptor inhibits endocytosis.";
RL Biochem. J. 267:31-35(1990).
CC -1- FUNCTION: CELLULAR UPTAKE OF IRON OCCURS VIA RECEPTOR-MEDIATED
CC ENDOCYTOSIS OF LIGAND-OCCUPIED TRANSFERRIN RECEPTOR.
CC -1- SUBUNIT: DIMER OF SIMILAR OR IDENTICAL CHAINS LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- PTM: EACH CHAIN BINDS GLYCANS, PHOSPHATE, AND A PALMITATE GROUP.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
DR EMBL: L19142; AAA03576.1;
KW Transmembrane; Glycoprotein; Receptor; Lipoprotein; Palmitate;
Signal-anchor; Endocytosis.

```

```

FT DOMAIN 1 65 CYTOPLASMIC.
FT TRANSMEM 66 89 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (BY SIMILARITY).
FT FT DOMAIN 90 757 EXTRACELLULAR:
FT FT CONTAINS TRANSFERRIN BINDING SITE
FT FT (BY SIMILARITY).
FT FT SIZE 20 23 ENDOCYTOSIS SIGNAL.
FT FT DISULFID 89 89 INTERCHAIN (BY SIMILARITY).
FT FT DISULFID 97 97 INTERCHAIN (BY SIMILARITY).
FT FT CARBOHYD 248 248 INTERCHAIN (BY SIMILARITY).
FT FT CARBOHYD 314 314 BY SIMILARITY.
FT FT CARBOHYD 719 719 BY SIMILARITY.
FT FT CARBOHYD 724 724 POTENTIAL.
FT FT CONFLICT 20 20 Y->T (IN REF. 2).
SQ SEQUENCE 757 AA; 85080 MW; A6A1749B CRC32;

Query Match 35.8%; Score 1504.5; DB 11; Length 757;
Best Local Similarity 42.8%; Pred. No. 3.9e-109;
Matches 334; Conservative 119; Mismatches 248; Indels 79; Gaps 19;

OY 35 EEEEDGEGAEETLAHFCPMELRGEPRLGSRPPNLIWMAAGRAAPYLVTALLIFT 94
Db 44 EENETDNMKASVRKH---RLNGRLCFGT-----IAVIFFL 78
OY 95 GAFLLGYAF--RSCQAC-----CDLVYVSEDVNTEPDDDFHGRGLYNSDQAMF 144
Db 79 TGFMTGYGYCKRKQKCVLAETGNSSEIIDEENP-----OSSRLYMDLKL 131
OY 145 LQFLGEGLEDITROTSLREVRVAGSAGMALTDODIRALSRQKLDHWTDTHYVGLFPD 204
Db 132 SEKDIAIEFTTITQLOSTSRAGSQDENLAYIENQFRFKLSKVRDHYVQKVG 191
OY 205 PAHPNTLHWVDACKVGEQLPLEDPDYVCYSAIGNVTGELVYAHYGRPEDLDLARGV 264
Db 192 SAAQNAVITIIINVN---GSDLVENPGGVAYASKATVSGKLIFANFGKFDLKY--- 246
OY 265 DPV--GRLLVAVGYISFAQKTNNAQDGAQVLIYEPADISQDPKPSLSSQAVYGVH 323
Db 246 -PVNGSLIVIRAGKITFEKAKVNAQSFNAIGVLIYMDTKF-----PVVAELSLFGHA 298
OY 324 HLGNGDPYTPGSPFNQGFPPVASSGSLPISAPDISADIRLRLKLVPAPOEMOGS 383
Db 299 HLGNGDPYTPGSPFNQGFPPVASSGSLPISAPDISADIRLRLKLVPAPOEMOGS 357
OY 384 LLSGPHLGPRLRLVNNHRTSTPINNIFGCIETRSEPDHYVIGARQAMGGAAS 443
Db 358 SL-CKLESSQGINVNLVYNNVLKETRILNIFGVIKGEFEPDRYIVVGAORDAMGGAAS 416
OY 444 AVGTALILELVYRTSSMWS--NGFRPRRLSLFISWDGDFGSGVSTEMLEGYLSVLHLKAV 502
Db 417 SVGTGLLKLKLAQFSDMVSRRGKFPKRSRIIPASVAGGEGAVGATEWLEGLSLHLKAF 476
OY 503 VYVSLDNAVGLGDKFNAKTSPLTSLIESVLKQVDSNPHSQCOTLYEQVFNPSMDAEVI 562
Db 477 TYINIDKVLVLTTRFKVSAAPLITLIEKTMQDVRRH- IDCKRPLXR-----DSMWISKV- 530
OY 563 RPLPMDSSAYSTAFVGVPAVEFSFEMDDQAYPFLTKEDITYENL-HKVLOGRLPAVAQ 621
Db 530 EDLSLDNAVGLGDKFNAKTSPLTSLIESVLKQVDSNPHSQCOTLYEQVFNPSMDAEVI 586
OY 622 VAQLAGOLLRLSHDRLLPLDFGRYGDVLRHIGNLNEFSDDLKARGITLQWYSARGDY 681
Db 587 AAEVAGQPIIKRLTHDIELNDYDNNKRLSFVELNFRADIRAMGSLQMLVSARDF 646
OY 682 IRAEKLQEIYSESEDERLTRMYNVRIMVEYFLSQYVSPADSPRHRIFMGGRDHTL 741
Db 647 FRATSRLLTTHDHNAAKTRFVYRELNNRIMKVEYHFLSPYSPESPRHRIFMGSGHTL 706
OY 742 GALLDHLRLRLNSNGSGTGATSSIGFQSRERRQALATLWTLOGAANALSGDVNMDNNEF 801
Db 707 TALVENLKLRLQKNSS-----AFNETLPRNQALATLWTLOGAANALSGDIVMDIDNEF 757

```

```

RESULT 3
ID 062351 PRELIMINARY: PRT: 763 AA.
AC 062351;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)
DE TRANSFERRIN RECEPTOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BLXDBA/2; TISSUE=HEMATOPOIETIC;
RA THOMBRIDGE I.S.; DOMINGO D.L.; THOMAS M.L.; CHAIN A.;
RU Submitted (JAN-1991) to the EMBL/Genbank/DBJ databases.
DR EMBL; X57349; CAA40624.1;
SC SEQUENCE 763 AA; 85731 MW; 5C2B158C CRC32;

Query Match 34.5%; Score 1450.5; DB 11; Length 763;
Best Local Similarity 41.7%; Pred. No. 6.5e-105;
Matches 333; Conservative 127; Mismatches 257; Indels 81; Gaps 24;

QY 24 ORVEPRRGHLE-----DEEDGEGATLAFHFCMELRGPPLGSRROPRLIPMAAG 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 27 ROVDGD-NSHVEKMLADEENADNMKA-----SVRKPRFRNGR-----LCFAA-- 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 79 RRAAPYLVTALLFTGAFLLGYVA-FRSGCOAGDS-----VLVSEDEVNEPDLDFH-- 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 71 -----IALVIF--FLIGFMSGYIGYCKRVQKECKYKLAETEETDSSEMETEDV 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 132 --QGRLYWSDIQAMFLQFLEGREEDTIRQTS--LRERVASGAMALTDIRALRSQ 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 119 PTTSLRYWADKTLSEKLNIEFADTIKQLSQNTYTPREGSQSDSLAYIENQFHEF 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 187 KLDHWYTHYVGOFPRAHPNTHLWYDEAKVGEQLPLEDPDYCYSAIGNTGELY 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 179 KFSKVRWDEHYKIOVKSIGSQNMVTIVOSNGND--PVEPEGYVAFSKPEVSGKLY 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 YAHGRPELDLDRARGVDPRLLVFRGVISFAQKTYNADDFAGVLLIPEPADSQ 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 236 HANGTKRDFEEL---SISVNSGLYIVRAGETTFEAKVAMNOSFANIGVLIYMDNKR-- 291
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 307 DPPKPSLSSQAAVYGVHVGTLGDPYTPGPFPSENOTQFPVYASSGLPSIPAOPISADIASR 366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 291 ---PVEADLALFGHAHLGTGDPYTPGPFPSEHNTQFPSPSSGLPNTIPVQISRAAEK 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 367 LLRKIKGVAAQEMOGSLGSPYHLGPGPRRLVLYNNHRTSPINNITGCIEGRSEPDHY 426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 347 LFGKEGGS-CPARMNID--SSCKLELSQNONMKLYKNVLEKRRILNIGVILGYDEPRXY 404
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 427 VVIGAQRAPWPG-AAKSAVGTAILLELVRTFSSMWS-NGFPRRSLEFISMDGDFGSV 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 405 VYVGGQRALAGVAAKSSVGTGLLKLQVYSDMISKDGFPPSSIIIPASTAAGDFGAY 464
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 485 GSTEMLEGYLSVHLKAAVYVSLDNAVLDGDKFNAKTSPLTSLIESYLKOYSDPNHSGQ 544
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 465 GATEWLEGYLSLHKAFYIYLDKVLGTSNFKVASAPLTYLGMKIMQDKHP-VDGK 523
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 545 TLIEVYVNTNSMDAEVIRPLPMDSSAISFTAFVGVPAVEFMDDDAIPPLHAKEDTY 604
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 524 SLYR-----DSNMISKY- EKLSFDNAAYPFLAYSGIIPAVSFECBDAD--YPLGTGRLDTY 576
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 605 ENL-HKVLQGRLPAAVQAVOAGOLLRLSHDRLLPLDFGYYGVVLRHIGNLEFSD 663
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 577 EALITKVPQ--LNMVRRARAEVAGOLIIKTLIDVELINDIYENKSLFKMDLQKFTD 634
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 664 LKARGLTIQWYSARGDYIRAEKLRQELISEERDEBLTMYNVRIMVEVEFLSQYVS 723
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 635 IRDMGLSIQWLYSARGDYFRATSRLLTDFHNAEKTRNFVMEIRIMKVEVHFLSPYVS 694
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 4
ID 099376 PRELIMINARY: PRT: 622 AA.
AC 099376;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE TRANSFERRIN RECEPTOR PROTEIN (TR) (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA MEDLINE: 91125359.
RA ROBERTS R.P.; GRISWOLD M.D.;
RT "Characterization of rat transferrin receptor cDNA: the regulation of
RT transferrin receptor mRNA in testes and in Sertoli cells in
RT culture."
RU Mol. Endocrinol. 4:531-542(1990).
CC -1- FUNCTION: CELLULAR UPTAKE OF IRON OCCURS VIA RECEPTOR-MEDIATED
CC ENDOCYTOSIS OF LIGAND-OCCUPIED TRANSFERRIN RECEPTOR.
CC -1- SUBUNIT: DIMER OF SIMILAR OR IDENTICAL CHAINS LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SIMILARITY: 88% SIMILARITY TO HUMAN TRANSFERRIN RECEPTOR
   (AC P02786).
DR EMBL; M58040; AAA42273.1;
RW Transmembrane; Glycoprotein; Receptor.
FT NON_TER 1
FT DOMAIN <1 622
   EXTRACELLULAR
   (CONTAINS TRANSFERRIN BINDING SITE)
   (BY SIMILARITY).
FT CARBOHYD 109 109
   POTENTIAL.
FT CARBOHYD 113 113
   POTENTIAL.
FT CARBOHYD 179 179
   POTENTIAL.
FT CARBOHYD 584 584
   POTENTIAL.
FT CARBOHYD 589 589
   POTENTIAL.
SC SEQUENCE 622 AA; 70152 MW; 7120562C CRC32;

Query Match 33.9%; Score 1424.5; DB 11; Length 622;
Best Local Similarity 46.5%; Pred. No. 5e-103;
Matches 304; Conservative 98; Mismatches 209; Indels 43; Gaps 15;

QY 155 DTIRTS---LRERVASGAMALTDIRALRSQKLDHWYTHYVGOFPRAHPNTL 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 DIKQLSQNTYTPREGSQKQDENLAIYENT.FHDFKSKVRWDEHYKTIQVKNYSYQN-L 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 212 HWYDEAGRVGEOLPLEDDVYCPYSALNVTGELYAHYHGRPELDLDRARGVDPVGRLL 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 64 VTINGSNID--PVEAEGVAVTSKAGEVYGLVHANFGKRDKEEL---NYSVNSGLV 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 272 LVRRGVISFAQKTYNADDFAGVLLIPEPADFSQDPKPSLSSQAAVYGVHVGTLGDPY 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 118 IYRAGKTIYFAEKVAMNOSFANIGVLIYMDRNTF-----PVEADLQFEGHAHLGTGDPY 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 332 TPGPFSNTOFPVASSGLPSIPAOPISADIASRLRKIKGVAPAQWQSGSLGSPYHL 391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 172 TPGPFSNHTQFPSPSSGLPSIPVQISRAPEKLFKNMGN-CPSPWNID-SSCKLEL 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 392 GPGPRRLVLYNNHRTSPINNIFGCIEGRSEPDHYVYVIGAQRDAPGAASAVGTAILL 451
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 230 SQONVVKLTIVNNVLEKETIILNIFGVIKYEEDPRYIVYGAQRDAMGPGVAAKSSVGTGLL 289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

Db 557 VYETVELVAFYD---PTEKHLTLVAQVBRRAWEFLANSIVLPEDQSAVALKKYADTI 613
QY 658 NEES-----GDLKARGTLQWVYSGARGDYITAAEKLREIYSSEERDRLTRMVRMRY 713
Db 614 YNISKHPOEMEAHMSFSDLSFSAVNNFTVASKFNGRLQELDKSNILLRINNDQMLYL 673
QY 714 EYFELSQYVPASPSPFHIFMGRGDH-----TLGALLDLHRLRLRSNSGTPGATSTSGF 767
Db 674 ERAFIDPLGLPGRPFRPHITTAAPSSHNKYGESFPGIYDLFDISKVNKNSKA----- 727
QY 768 QESRFRRLALLTWTLQGAANAL 790
Db 727 -WSEVKROIATETVQAAETL 748

RESULT 10
ID 077564 PRELIMINARY; PRT: 751 AA.
AC 077564;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE FOLYPOLY-GAMMA-GLUTAMATE CARBOXYPEPTIDASE.
US Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=JEJUNAL MUCOSA;
RX MEDLINE: 98352082.
RA HALSTED C.H., LING E., LUTHI-CARTER R., VILLANUEVA J.A., GARDNER J.M.,
RA COYLE J.T.;
RT "Folypoly gamma-glutamate carboxypeptidase from pig jejunum.
RT Molecular characterization and relation to glutamate carboxypeptidase
RT II."
RL J. Biol. Chem. 273:20417-20424(1998).
DR EMBL: AF050502; AAC39269.1; -.
KW Carboxypeptidase.
SQ SEQUENCE 751 AA; 84523 MW; 0F6E680A CRC32;

Query Match 16.9%; Score 709.5; DB 6; Length 751;
Best local Similarity 26.9%; Pred. No. 5e-47;
Matches 212; Conservative 144; Mismatches 314; Indels 117; Gaps 25;

QY 65 RPRQPLIPMAAAGRRAAPLVLTALITFGAFLGVAFRGSCQACDSVLVYSEDPVY 124
Db 17 RPR-----WLCAG-----ALVLAAGLFLVGL-----FGWFIKSPNEANL 52
QY 125 EPRDLFHQGLVMSDLOAMFLQFGEERLEDITRQTSLSREVRVAGSAGMALTDIRALSL 184
Db 53 SPQ-----HNWKAFLDELKAKENITFTFYNFRIPLHLAGTEONFOLAKIOQOMK 102
QY 185 ROKLDHWYTDTHYVGLQEPRAHPNTHLWVDEAGKVGEOPLDEP-----DVCY 235
Db 103 EFLGDSNLAHYVLSYPRKTRPNYISIIDEGNEIFNTSLPEPPPGYENNSDVVPF 162
QY 236 SAI---GNVTGELVYAHYGRPEDLODL-RARGVDPVGRLLLVRYGVISFAQKVTNAODFG 291
Db 163 SAFSPQGMPEGLVYVYARTDEDFEKERDMKINCSGKILIAARKGLFRGNKYKNAQLAG 222
QY 292 AOGVLIYPERADF-----SDDPRPSLSQQAAYGVH-HL-GTGDPTTPEGPSPNQTQFP 344
Db 223 AKGILISDADYFAPGVOSYPRGWNLPGGGVORGNITLNGAGDPLTPGYPAANEVAYRL 282
QY 345 PVASS-GLPSIPAPDISADLASRLRLKLGKPVAPQF-WQSGILGSPYHLGPG-----P 395
Db 283 QINAEVGLRPIYHPRTISDAOKLLEKMGSAFPDDSKKS-LHVPYNWVGEGFIGNSTQ 341
QY 396 RLRLVNNHRTSPFINNIFSCIEGRSEPDHYVIGAQRDAMGPGAAKSAVGTALILELVR 455
Db 342 KYKMHINSDMKVKRYIVNVIGTLRGAVERPDRYVILGHRDSWVFEGIDPOGSAAVVHIVR 401

```

```

QY 456 TFSWVSNGRPRRSLFTSMQGDGFSVSGSTEWTEGLYSLVLRKAVVYSLDNVIGDD 515
Db 402 SFGKLKKEGMRPRRTVLFASWDAEEYGLFGSTWAEENSRIQERKVAATNADSIIGNV 461
QY 516 KFAKTSPLTSLIESVLQAVDSPNH--SGQTLVEQVFTNPMDAEVIRP-----LPMD 569
Db 462 TLAVDCTPLMYSLVNLTRELQSPDGEFEKSLFE-----SWNEKSPSESGLPRI 514
QY 570 ---SAVSFAF---VGVPAVESFMEF-----DQAVPLHTKEDTYENLHKVLQGLRPAY 618
Db 515 KLCSGNDVEFEVFORLGIASGRARYTRKDWYTKFSSYPLXHSVYETVELVEKFDY---PTF 571
QY 619 AQ--AAVQLAGOLLIRLSHRLPLDGRYGDVVLHIGLNFEF---GDLARGLITLQ 672
Db 572 KYHLAAVQVAGTIVFELANSVPRPCRDYAVVLRVADKLXNISMNHPQEMAYSVSF 631
QY 673 WYVSARGDYITRAAEKLRQEIYSSEERDERLTRMYNVRIMRVEPFYFQVYSPSPRH 732
Db 632 SLFSAKNFTFELANSFNSERQDLDKNNPILLRINMQMLHFAFIYPLGLPBRATYRHV 691
QY 733 EMGRGDHTLQALLDLRLRSNSGTPGATSTGFOESR-----FRQALALLTWTL 783
Db 692 IYAPSSHN-----KMGESFPGIYDALFDIENKVDPSKANGVEYKQISIAAFV 740
QY 784 QGAANAL 790
Db 741 QAAAGTL 747

RESULT 11
ID P91406 PRELIMINARY; PRT: 751 AA.
AC P91406;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE SIMILAR TO HUMAN PROSTRATE-SPECIFIC MEMBRANE ANTIGEN.
GN R57.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEERS J.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIKKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOHLDMAN P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA FAVELLO T., RIKIN L., CHIAPELLI B.;
RA Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RA Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RA Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.

```

DR EMBL: U88179; AAB52660.1; -
SQ SEQUENCE 751 AA; 83663 MW; 088B6F82 CRC32;

Query Match 14.8%; Score 621; DB 5; Length 751;
Best Local Similarity 27.1%; Pred. No. 4,1e-40;
Matches 177; Conservative 126; Mismatches 283; Indels 68; Gaps 18;

OY 151 GLELETIQTSIRE-----RVAGSAGMALIOTDIALSROKLDHVTMDTHVGLQ 201
DB 47 GVIQNIKIKETIKKRLFTKDPHVAETANKKVAEELIANNMSEAGLEVHLPLEVLLS 106
OY 202 FDPDPHPTLHWVDAGKVGEO-----LPLEDPDVYC-----PSYAGNVTGELYA 248
DB 107 YDFENPNVSIITSSAGKVEFKSKVSPYITIPDEOSGKTAGHOMLALYAGSASADVYI 166
OY 249 HXGREDDLODLRARGVDPVGRLLLVVGVISFAQKVTNAODFGAOGVLIYEPADFSOD- 308
DB 167 NGGTANDEFKNTLKGVDIKIALMRVHGFRGDKIKRQQAQALIGALLIFSDTDVAODG 226
OY 308 -----PPKPSLSQQAIVYGHVHGTGDPYTPGFPSS-----FNQTOFPVVASG-LPSTP 355
DB 227 VESENVYPPKKIWMPEGVQVQSLMHGDDALSPYPSKKELEKRTIEBAKEDGVLPSP 286
OY 356 AOPISADIASRLRLKRGVPAPQEMOGSLGS-PYHLGP-----GPRRLVYNNHRTSP 410
DB 287 VLPVYITTCYELIKRLSGRPAPSDMOGFVGNLYTKLGFVNGEKLSINVSELETKRTI 346
OY 411 NNFICIEGRSEPDHVVIGAQORDAMGPAKSAVGTALILEVTFSSMVS-NGFRPRR 469
DB 347 RVVIGYIRSEEPDSYIMGNHFDAMVYGSIDPNSGTAVLAEVARAMQOTINETSMKPAR 406
OY 470 SLFLTSMOGDGSVSTFMLEGYLSVLKAVVYVSDNNAVLGDDFKAKSPILTSI 529
DB 407 TIVFANMADEEFGLSIEEVEFNILQKRAVYIIMD-CIQGINSIHVDVPLEHNAV 465
OY 530 ESVLQVDSPP-----NHSQGLYE--QVFTNPMDAEVIRPLPMDSSAVSFTAEGVPA 582
DB 466 IASAGVEMPSPKRSRSGKTIYDTMMKYFPKKKAGVPRIRPVGSGSDAPLNFAGVY 525
OY 583 VEFSEMEDD--QAYPPLTKEDT-YENLHKVLOGRLPAVAQAVALAGOLLIRLSHRL 639
DB 526 INFETKNTYMDTYPLVHMYETPFPSNILLDTDNL-SVHKAIGQYMALEAKTAPADVLT 584
OY 640 PLDPGRYGDVLR-----HIGNLNEPSGDLKARGTLQWVYASRGYITAAEKLRQ 691
DB 585 PNMTHFASVMLKTYLPOLKTTISGINVSRSDPEDIRTOYALLSKSAODLLMSKRFQET 644
OY 692 I-----YSEERDERLTRMYNVRIMRVEFYLS-QVSPADSPFHHIMGSD 738
DB 645 IFTQHSFQNPYDPKHNVAVERLKTSTRCFINRGVSMHNP SARHVLFSVSD 698
RESULT 12
C43023 PRELIMINARY; PRT; 794 AA.
AC 043023:
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-AUG-1998 (TReMBLrel. 07, Last annotation update)
DE HYPOTHEICAL 89.0 KD PROTEIN.
GN SPAC354.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;
RA WOOD V., RAJANDREAN M.A., BARRELL B.G., HILBERT H., DUESTERHOEFF A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022071; CAA17809.1; -
KW Hypothetical protein.

SQ SEQUENCE 794 AA; 88965 MW; 6A5AA0A9 CRC32;

Query Match 12.9%; Score 541; DB 3; Length 794;
Best Local Similarity 26.3%; Pred. No. 8e-34;
Matches 192; Conservative 121; Mismatches 273; Indels 144; Gaps 26;

OY 84 YLVLTALLIFGAPFLIGVARG-----SCQACGDSVLVSEVNYEPEDLDFHGRGLYMS 138
DB 97 YLAVATATVITLILASY--YGFDPGVAMRSVHHYSEDVL-----LNTIKGCC----- 140
OY 139 DLQAMFLOFLGEGRLIEDIRQOT-----SLREVRVAGSAGMALIOTDIALSROKLDHVTMDT 195
DB 140 -----ISDRQOVMTLSSIPHLAGVIGSSSLQIMNRLRYKRTITVDPRE 185
OY 196 HVGLOFP-----DPAHPNTLHWVDAGKVGEOPLLEDPDVYCPSYAGNVTGEL 245
DB 186 FYAVLNFQVLVSLSDIDGDSFHPSLIESYQGVGVDSVLP--TPATF-GGSPSGFVANPL 242
OY 246 VYAHYGREDDLODLRARGVDPVGRLLLVVGVISFAQKVTNAODFGAOGVLIYEPADFS 305
DB 243 VYANRGRIEDEEMLVNSGIYESSIVLVRANQSDFALATANAKEYNNAALIFEDTYLTS 302
OY 306, OD-----PPKPSLS--QQAIVYGHVHGTGDPYTPGFPSSNOT-QFPVASSGLPSI 354
DB 303 LDNLNQVYPAPRYPSANSLYRGSVANHYT-VGDPILPQWNAHETNRISPKDANVLPST 361
OY 355 PAOPISADIASRLRLKRG-----PVAPQEMOGSLGSPYHLGPRLRLT 399
DB 362 VSIPTFNDGIELLRLQGHGLVKNWCODLAPVLEVTGSKISS-----PGLVYN- 416
OY 400 VYNNHRTSPINNIFGCIEGRESEPDHVVIGAQORDAMGPAKSAVGTALILEVTFSS 459
DB 416 VLQDIEDKQKTIINMAQIDG-YESDQILVWGPAPDSWCTGASDSSVSTLLIDVISTPAN 474
OY 460 MYSN-GFPRRSLLFTSMWGDGFSVSTFMLEGYLSVLKAVVYVSDNNAVLGDDFKH 518
DB 475 MAQDLSMKPRRTIYFASMDARQFNAISTEFLVWKESLEKAAVAINVDVAVG-DFTT 533
OY 519 AKTSPLTSLIESVLKQVDSPNHSQGLTYEQVFTNPMDAEVIRPLPMDSSAVSFTAEGV 578
DB 534 ARTVGLKRVQI---RAFDVANEDEMKRANIIINDDEFTYSDLT-----SFLTFA 580
OY 579 GVPANVEFSFEMDDQ--AYPPLTKEDT-----YENLHKVLOGRLPAVAQAVAL 624
DB 581 GIPVYNLAERNEENPPMPFLGSCEDTVSWIDTFGSEYWN-----AARLGK 628
OY 625 LAGOLLIRLSHRLPLDPFRYGDVLRHIGNLNEPSGDLKARGTLQWVYASRGYITRA 684
DB 629 IWSYILLFLANDPVVYPLEDEINGVGMKRIPIRP--ANNLDRKINEESELLS 685
OY 685 AEKLRQEI-----YSEERDERLTRMYNVRIMRVEFYLSQVSPADSPFHHIF 733
DB 686 LIRFEDEIREWKSIMMNSYTVSVKHPLELGYNAKLARFERSFLDEAGLGHMYKHLI 745
OY 734 MG---RGDHT 740
DB 746 YGPNLRNSHS 755
RESULT 13
093332
AC 093332 PRELIMINARY; PRT; 1483 AA.
AC 093332:
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
DE C35C5.2 PROTEIN.
GN C35C5.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]

ID 061560 PRELIMINARY; PRT: 277 AA.
AC 061560:
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-NOV-1998 (TRENBLREL. 08, Last annotation update)
DE TRANSFERRIN RECEPTOR (FRAGMENT).
GN TRFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85159078.
RA STEARNE P.A., PIETERSZ G.A., GODING J.W.:
RT "cDNA cloning of the murine transferrin receptor: sequence of
RT trans-membrane and adjacent regions."
RL J. Immunol. 134:3474-3479(1985).
DR EMBL: M29618; AAA37616.1; -.
DR MGD: MGI:98822; TRFR.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 277 AA; 31207 MW; EBE0C8AE CRC32;

Query Match 6.5%; Score 271.5; DB 11; Length 277;
Best Local Similarity 27.9%; Pred. No. 1,7e-13;
Matches 88; Conservative 54; Mismatches 116; Indels 57; Gaps 13;

QY 24 QNVEGPRKGLHLE---EEEDGEGEAGTIAHFCPEMLGPEPLGSRPRQPNLIIPWAAG 78
DB 3 ROVDGD-NSHVEWKLADEENADNMKA-----SVRKPFRNGR-----LCFAA-- 47
QY 79 RRAAPYLVTALLITGAFILGYVA-FRSCQACGDS---VLVSEEDVNYEPDLDFH-- 132
DB 47 -----IALVTF--FLIGMSGYLGCKRVEQKECVKLAETETDKSETMETEDV 94
QY 132 --QRLYMSDLQAMFLOFLGEGRLIEDTIR---QTSLEFRVAGSAGMAALTODIRALSRQ 186
DB 95 PTSSRIYVADLTLSSEKINSIEFADTIKQLQONTYTPREAGSQKDESLAYIENGFHEF 154
QY 187 KLDHWVTDTHYVGLQFPDPAHPNLTLMVDAGKVGEOLEPLEDDPVYCPYSAIGNVTGLV 246
DB 155 KFSKWRDEHYKIQYKSSIGQNMVTIVQSNGLD---PVESPEGYVAFSKPTEVSGKLY 211
QY 247 YAHYGRPELDQLRANGVPVGRLLLYRVGYISFAKQVTNADFGAGVLIYEPADFSQ 306
DB 212 HANFGTKDFEEL--SYSVNGSLVIVRAGEITFAEKVANAQSFNAIGVLIYDNKRF-- 267
QY 307 DPPKPSLSQQAAYG 321
DB 267 ----PVVENDLALFG 277

Search completed: January 13, 2000, 12:12:05
Job time: 103 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 13, 2000, 08:36:16 ; Search time 1456.15 Seconds

(without alignments)
6772.760 Million cell updates/sec

Title: US-09-358-755-2

Perfect score: 2877

Sequence: 1 ctgcacggctcagggaggga.....ctgatacataaagccaanaa 2877

Scoring table: IDENTITY_NDC

Searched: 4089388 seqs, 1713965092 residues

Database: EST.*

Word size: 0

Number of hits that pass the threshold : 8178776

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gp_est1:*
21: gp_est2:*
22: gp_est3:*
23: gp_est4:*
24: gp_est5:*
25: gp_est6:*
26: gp_est7:*
27: gp_est8:*
28: gp_est9:*
29: gp_est10:*
30: gp_est11:*
31: gp_est12:*
32: gp_est13:*
33: gp_est14:*
34: gp_est15:*
35: gp_est16:*
36: gp_est17:*
37: gp_est18:*
38: gp_est19:*
39: gp_est20:*
40: gp_est21:*
41: gp_est22:*
42: gp_est23:*
43: gp_est24:*
44: gp_est25:*
45: gp_est26:*
46: gp_est27:*
47: gp_est28:*
48: gp_est29:*
49: gp_est30:*

50: gp_est31:*
51: gp_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gp_est33:*
60: gp_est34:*
61: gp_est35:*
62: gp_est36:*
63: gp_est37:*
64: gp_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gp_gss1:*
70: gp_gss2:*
71: gp_gss3:*
72: gp_gss4:*
73: em_gss1:*
74: em_gss2:*
75: em_gss3:*
76: em_gss4:*
77: gp_gss5:*
78: gp_gss6:*
79: gp_gss7:*
80: gp_gss8:*
81: gp_gss9:*
82: em_gss5:*
83: em_gss6:*
84: em_gss7:*
85: em_gss8:*
86: em_gss9:*
87: em_gss10:*
88: em_gss11:*
89: gp_gss10:*
90: gp_gss11:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	542.8	18.9	548	61	A1814648	A1814648 wj75f09.x
C 2	520	18.1	536	38	AA748421	AA748421 nyo1a04.s
C 3	481.4	16.7	700	50	AJ7225643	AJ7225643 RNJ225643
4	443.6	15.4	504	27	AA007639	AA007639 zh98d08.r
5	414.4	14.4	782	43	AU035332	AU035332 AU035332
C 6	393.2	13.7	452	23	R99295	R99295 yq71d03.s1
C 7	388	13.5	445	22	R46386	R46386 yq50d07.s1
C 8	382	13.3	425	24	R49459	R49459 yq68f02.s1
C 9	371.8	12.9	444	27	AA007614	AA007614 zh99d08.s
C 10	368.8	12.8	617	41	A1047358	A1047358 ud65d02.y
C 11	361.4	12.6	384	34	AA505641	AA505641 nh82a03.s
C 12	351	12.2	388	37	AA682374	AA682374 zj86a08.s
C 13	348.6	12.1	363	38	AA809686	AA809686 n299d04.s
C 14	334.6	11.6	384	24	H63101	H63101 yf50h04.s1
C 15	332.4	11.3	366	24	H70432	H70432 yf03c12.s1
C 16	324.8	11.3	503	22	R35943	R35943 yq68f02.r1
C 17	318.2	11.1	341	37	AA693893	AA693893 z149d07.s
C 18	299.6	10.4	367	26	W89178	W89178 zh74h11.s1
C 19	292.6	10.2	708	41	A1046365	A1046365 ud65d02.x
C 20	273	9.5	506	49	A1663104	A1663104 uk22g09.y
C 21	267.4	9.3	408	21	T64983	T64983 yd11e08.s1
C 22	266.2	9.3	323	21	T98167	T98167 yq56c10.s1

23	252.4	8.8	581	35	AA537969	AA537969 vj36d09.r
24	236.2	8.2	421	45	AI386238	AI386238 mm20c12.Y
25	235.2	8.2	264	21	T98166	T98166 y636c10.r1
26	225.8	7.8	465	21	T72291	T72291 yc68e08.r1
27	224	7.8	225	38	AA776476	AA776476 z172h01.s
28	220.6	7.7	384	48	AI551005	AI551005 vj28f05.Y
29	220.4	7.7	884	51	AU050369	AU050369 AU050369
30	220.2	7.7	1000	51	CB1942	CB1942 CB1942 Leuk
31	214.8	7.5	269	24	HE3187	HE3187 yf50h05.r1
32	211	7.3	270	26	W89177	W89177 zh74h11.r1
33	208.4	7.2	675	48	AI596094	AI596094 uk22g09.x
34	203.2	7.1	655	44	AI256087	AI256087 u194h08.x
35	203	7.1	207	22	R29632	R29632 F1-1153D.22
36	203	7.1	436	51	AU050545	AU050545 AU050545
37	199.2	6.9	234	31	AA511579	AA511579 vj28f05.r
38	198	6.9	234	21	T87554	T87554 yd89h12.s1
39	190.8	6.6	687	43	AI132639	AI132639 mm20c12.x
40	184.6	6.4	461	43	AI195082	AI195082 u160d07.x
41	157.2	5.5	457	41	AI019669	AI019669 ua92d01.r
42	136.2	4.7	458	64	AL120908	AL120908 DKRP2762P
43	116.4	4.0	699	51	CB1968	CB1968 CB1968 Leuk
44	112.4	3.9	187	28	AA096819	AA096819 mm20c12.r
45	101.4	3.5	106	24	H70431	H70431 ys03c12.r1

ALIGNMENTS

RESULT 1
LOCUS AI814648 548 bp mRNA EST 24-AUG-1999
DEFINITION w75f09.x1 NCI CGAP.Lu19 Homo sapiens cDNA clone IMAGE:2408681 3'
similar to PR:075422 075422 TRANSFERIN-RECEPTOR2.1; mRNA

ACCESSION AI814648
NID 93425863
VERSION AI814648.1 GI:5425863
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 548)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187613.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskajuk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LENI at:
www-bio.llnl.gov/bdrrp/image/image.html

FEATURES
Seq primer: -40UP from GIDCO
High quality sequence stop: 460.
Location/Qualifiers
1..348

Source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2408681"
/clone_lib="NCI-CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"

/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pTR3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dt) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pTR3 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 126 a 135 c 165 g 120 t 2 others
ORIGIN

Query Match 18.9%; Score 542.8; DB 61; Length 548;
Best local similarity 99.3%; Pred. No. 1,4e-97;
Matches 544; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2326 ttccagagagccgttccgcgcgtcagctagccctgtctacctgagcgtgcaaggcca 2385
548 TTCCAGGAGAGCCGTTTCCGGCGTCACGTACCCCTGCTCACCTGAGCGTCAAGGGGCA 489
2386 gccaatgcgtctagcgggagatgctcgaacattatatacaacctctgagggccctgggat 2445
488 GCCAATGCGCTTAGCGGGGAGTCTGCAACATTATACAACTCTGAGGCGCCTGGGAT 429
QY 2446 cctacatcccgctcccccagtcgaagctccctctgctcgtctgaatgattcaaggt 2505
428 CCTCACATCCCGCTCCCGCCCACTCAAGAGCTCCTGCTCCGCTTAATGATTCAGGGT 369
QY 2506 caggagagtgctcgaagctcgaagctcgaagctcgaagctcgaagctcgaagctcgaag 2565
368 CAGGAGAGTGCTCAGAGTCACTCCTCATGCTGATGATCAATTTCTCATTTACCTTACACA 309
QY 2566 tctctcagagagccagaccagaccagacagatataccacaccccccagctgagctgag 2625
308 TCTCTCCAGCGAGCCAGCCAGCCAGCAGATATCCACACCCCGCCTGCGATGATG 249
QY 2626 ctgacacctaatgtgacggtatctactgtcggttaactcagaagagtagcatccctaatcac 2685
248 TTGACACNTAATGATGACGCTCATCTGCTGCTAATCAAGAGTGAATCCCTTCAATCAC 189
QY 2686 agcccttcccttctcggggtctcctacatcctagagacacctctgggaggttgctaaag 2745
188 AGCCCTTCCCTTCTTCTGCGGCTCTCATCTAGAGACACTGTGGAGGTTTGCTAGG 129
QY 2746 cccctggagacctggcgaacctctgttagtggagagagatcgtggaaccataagcctatggcc 2805
128 CCGTGGAGCCTGGCCACGCTGTAGTGGGAGATGCGTGGCACCATPAGGCTTATGGCC 69
QY 2806 aacaggtggtctgtgtgaaagggtggtgagttcaatatcaataaacacacctgatatc 2865
68 AACAGGTGCTGTGTGTAAGGAGGCGTGGATTTCATATCAATTAACACCGATATTC 9
QY 2866 aataagcc 2873
Db 8 AATAAGCC 1
RESULT 2
LOCUS AA748421/c 536 bp mRNA EST 18-FEB-1998
DEFINITION ny01d04.s1 NCI-CGAP.GCB1 Homo sapiens cDNA clone IMAGE:1270446 3'
similar to SW:TRSR_HUMAN P02786 TRANSFERIN RECEPTOR PROTEIN ;,
mRNA sequence.
ACCESSION AA748421
NID 92788379
VERSION AA748421.1 GI:2788379
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 536)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1798396.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
www-bio.11nl.gov/bdrrp/image/image.html

Insert Length: 1516 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 385.
 Location/Qualifiers
 1..536
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="1"
 /clone_1id="NCI_CGAP_GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="vector: p773D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). CDNA synthesis was
 primed with a Not I - oligo(dt) primer
 [5'-TGTTACCAATCTAAGTGGAGCGCCCTCATTTTGTGTTTGTGTTT-
 3']. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p773 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 119 a 125 c 160 g 132 t
 ORIGIN

Query Match 18.1%; Score 520; DB 38; Length 536;
 Best Local Similarity 99.1%; Pred. No. 4,4e-93;
 Matches 523; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2350 cagctagccctgcctcacctgagcgtgcaaggaggcagcgaatgctgtaggggagatgc 2409
 |||||||
 DB 536 CAGCTAGCCCTGCCTCACCTGAGCGCTGCAAGGCCAGCAATGCGCTTAGCGGGAGATGC 477
 |||||||

OY 2410 tggacaattgtaacaacttggagccctggggatccctcacatccctccctcagtc 2469
 |||||||
 DB 476 TGGACAATTGTAACAACCTCTGAGGCCCTGGGGATCCTCAATCCCGTCCCAAGTCA 417
 |||||||

OY 2470 agagctctctgcctcctgcctgctgaatgattcagggtcaggaggtgctcagatccacc 2529
 |||||||
 DB 416 AGAGCTCTCTGCTCCCTGCTGTAATGATTCAGGAGGTGCTCAGAGTCCAC 357
 |||||||

OY 2530 tctcattgctgataattctcattacccctcacacatctcccaaggagcccaagccca 2589
 |||||||
 DB 356 TCTCATTTGCTATCAATTTCTCATTTACCCCTACACATTTCTCCAGGAGCCCAAGCCCA 297
 |||||||

OY 2590 gcacagatatcacacacccagccctgcaagttagctgacccctaattgtagggtcatac 2649
 |||||||
 DB 296 GCACAGATATCACACACCCAGCCCTGCAGTGATGACCCCTAATGTGACGGGTCAATAC 237
 |||||||

OY 2650 tctcggttaatcagagagtagatcattcccttcattcacaccccttccttcctgaggtcc 2709
 |||||||
 DB 236 TGTGCGTTAATCAGAGAGTGTGATCCTTCAATCACAGCCCTTCCCTTTCTTGGGTCC 177
 |||||||

OY 2710 tccatacttagagaccactctgggaaggtttgctaaagccctggagcctgagcactctgt 2769
 |||||||
 DB 176 TCCATACCTTAGAGACCACTCTGGAGGTTGCTAAGCCCTTGGAGCCTGGCAGCTCTGTT 117
 |||||||

OY 2770 agtggagagatgcgtctgcaccataagccttaaggccaacaggtgtctgtgtgaaagg 2829
 |||||||
 DB 116 AGTGGAGAGATGCTGTGGCACCATTAGCCTTATGGCCCAACAGGTGCTGTGTGAAGGG 57
 |||||||

OY 2830 gcctgaggttcattcatcattaaacccctgatatcatgaagccaana 2877
 |||||||
 DB 56 GCGTGGAGTTTCAATATCAATTAACCACTGATATCAATATTAATAA 9
 |||||||

RESULT 3
 AJ225643
 LOCUS RNU225643 700 bp mRNA EST 14-MAY-1999
 DEFINITION RNU225643 Rat liver ESTs (E.Olivier) Rattus norvegicus cDNA clone
 IRL413, mRNA sequence.
 ACCESSION AJ225643
 NID 94833473
 VERSION AJ225643.2 GI:4833473
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutharia; Rodentia; Sciurognathu; Muridae; Murinae; Rattus.
 1 (bases 1 to 700)
 AUTHORS Olivier E., Soury E., Risler J.L., Smith F., Schneider K.,
 Lochner K., Jouzeau J.Y., Fey G., and Saller J.P.
 TITLE A novel set of hepatic mRNAs preferentially expressed during an
 acute inflammation in rat represents mostly intracellular proteins
 JOURNAL Genomics 57 (3), 352-364 (1999)
 MEDLINE 99263497
 COMMENT On May 6, 1998 this sequence version replaced gi:3114939.

CONTACT: E. Olivier
 U78 INSERM
 543 chemin de la Breteque, 76233 Bois-Guillaume, France
 library construction: J.B.C. vol.270, p29998-30006 (1995)
 POLYA-No.

FEATURES
 source
 1..700
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone_1id="IRL413"
 /clone_lib="Rat liver ESTs (E.Olivier)"
 /tissue_type="liver"
 /dev_stage="adult"
 /note="Organ: liver; Vector: lambda ZAPrii; Library
 construction: Ripberger, J.A. et al. J. Biol. Chem. (1995)
 270(50):29998-30006"

BASE COUNT 149 a 198 c 210 g 143 t
 ORIGIN

Query Match 16.7%; Score 481.4; DB 50; Length 700;
 Best Local Similarity 84.8%; Pred. No. 2e-85;
 Matches 598; Conservative 0; Mismatches 101; Indels 6; Gaps 5;

OY 1464 tgaatttggagagctgggcccacagagatggtcgaaggttacctgaagctgtgcaacct 1523
 |||||||
 DB 1 TGACTTTGGCAGCTGGGAGCCACGAGATGG-16GAGGCTACTCTGACGCTGCTACACT 59
 |||||||

OY 1524 caaagccgtagtgtacgtgagcctggaacagcagtgctggtggagatgacaaagtgtcagc 1583
 |||||||
 DB 60 CAAAGCCTGTGCTGTATGTAGCGCTGACCAATCCGTTTGGAGAGCGCAATTCATGTC 119
 |||||||

OY 1584 caagaccagccccccttgacaagctcattgagagtgctcctgaagcagtgatctcc 1643
 |||||||

LOCUS	AU035332	782 bp	mRNA	EST	08-OCT-1998
DEFINITION	AU035332 Sugano mouse brain mncb Mus musculus cDNA clone MNCB-0333				
ACCESSION	AU035332				
NID	93718335				
VERSION	AU035332.1	GI:3718335			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 782)				
JOURNAL	Sasaki,Z., Suzuki,Y., Watanabe,M., Imai,J., Shibui,A., Yoshida,K.,				
COMMENT	Hata,H., Yamaguchi,K., Tateyama,S. and Sugano,S. Construction of mouse full length-enriched cDNA libraries unpublished (1998) On Oct 17, 1997 this sequence version replaced gi:2518041.				
FEATURES	Contact: Katsuyuki Hashimoto Division of Genetic Resources National Institute of Infectious Diseases 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan Email: khash@nih.go.jp.				
SOURCE	Location/Qualifiers				
	1..782				
	/organism="Mus musculus"				
	/strain="C57BL"				
	/db_xref="taxon:10090"				
	/clone="MNCB-0333"				
	/clone_lib="Sugano mouse brain mncb"				
	/sex="female"				
	/dev_stage="adult"				
	/lab_host="TOP10"				
	/note="Organ: brain; Vector: pME185-FL3; 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTTCCTTTTTTTT] double-stranded cDNA was ligated to a DraIII adaptor [GTTGGCTACTGG], digested and cloned into distinct DraIII sites of the pME185-FL3. XhoI sites just outside the DraIII sites can be used to excise fragments <1.5 kb. Library was constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer for sequencing: 5' end primer [CTTGCTCTTAAGCTGCg]"				
BASE COUNT	170 a 235 c 208 g 157 t 12 others				
ORIGIN					
	Query Match 14.4%; Score 414.4; DB 43; Length 782; Best Local Similarity 78.2%; Pred. No. 3.3e-72; Matches 505; Conservative 0; Pmismatches 139; Indels 2; Gaps 1				
OY	642 gagctaccaccaacccctcactgttgatgatagggcggaagtgcaggagcagctgcc 701				
DB	39 ggctaacgttaaacaccttcgactgggtgatgcacagcgagagcgtcacagacgctaac 98				
OY	702 gctggaggaacctgaagcttactgccccctacaagcgcatcgcaagctgaagagct 761				
DB	99 gctgagagattcggagaagtactgtgtcctccacacgccaccggcmaagccacggcgaact 158				
OY	762 ggtgtacgccaactacggcgccgcgcgaagaactgcagaagcttcgagcgagggcgtgga 821				
DB	159 ggtttacgcccactacgtacggcggtgcgagagccttacagagacttaaaaggccaaaggcgtgga 218				
OY	822 tcgaatgggcgcgcgttgtgtgtgtggcggtggatcatcaacttcggcccaaagtgtac 881				
DB	219 gctggccggcaccctctcctctaagtgcgaatttggaatttacttactctgccacagaagtac 278				
OY	882 caatgctcagaagcttcggagcccaaagatgctcatatatccagaagcgagcgacttc 941				
DB	279 cgttgcccaagagactttggggcgtcacaagagatgctganatnacctgacacacttcac 338				
OY	942 ccaggaaccacccaagccaagcctgtccagccagcagcgagtgatagacatgttgacct 1001				

	DB	339	CCAGATCCCCCAACGAGCGCTTCTTAGCCACCAGGCTGTGTACGGACAATGTGCACCT	398
OY	1002	gggaactgagagacccctcacacacgtgcgtccttccttccttcaatacaaccagttccctc	1061	
DB	399	GGAACCTGGAGACCTTTACACACCTGGCTTCCTCCGCTCTTAATCAAAACCATTCCTTC	458	
OY	1062	agttcatcatcaagccttcccagatcccaagccccagccatcaagtgcacagattgctc	1121	
DB	459	AGTAGAATCATCAGGCGCTTCAGCATGCCCGCCAGCCCATTCAGTGCGTGCATGTGCTGA	518	
OY	1122	ccgcttgctgagaagctccaaggccttgtgcccaccaagaatgycagggagagctcct	1181	
DB	519	TCAATTGCTCAGAGAACATCACAGGCCCGGTGCTTCCAGGANTGAAAGGCGACTNTN	578	
OY	1182	agggccccccttaacacccgggcccgccagcaagctggcgtagtgcagtgatcaaatccag	1241	
DB	579	AGGCTCTMCTTATTMGCTGGGACCTGGGGCCGACATTACGCTTGAGGGCAACACCAACC	638	
OY	1242	ga-ccctcaccccccatcaacaacatcttcgctgcctcatogaagacc	1285	
DB	639	GAGCTMTTACCCCATTAGTAACTTTTGGCTGCATTGANGGCC	684	
RESULT	6			
R99295/c		R99295	452 bp	mRNA EST 14-SEP-1995
LOCUS		YC71d03.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone		
DEFINITION		IMAGE:201221 3', mRNA sequence.		
ACCESSION		R99295		
NID		9985896		
VERSION		R99295.1	GI:985896	
KEYWORDS		EST.		
SOURCE		human.		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS		Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
		1 (bases 1 to 452)		
		Hillier,L., Clark,N., Dubque,T., Elliston,K., Hawkins,M.,		
		Holtman,M., Holtman,M., Kucaba,T., Le.M., Lennon,G., Marra,M.,		
		Parsons,J., Rifkin,B., Rohlfing,T., Soares,M., Tan,F.,		
		Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and		
		Wilson,R.		
TITLE		The WashU-Merck EST Project		
JOURNAL		Unpublished (1995)		
COMMENT		On Nov 29, 1993 this sequence version replaced gi:430583.		
		Contact: Wilson RK		
		Washington University School of Medicine		
		4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
		Tel: 314 286 1800		
		Fax: 314 286 1810		
		Email: est@watson.wustl.edu		
		Insert Size: 1674		
		High quality sequence stops: 262		
		Source: IMAGE Consortium, LNL		
		This clone is available royalty-free through LNL; contact the		
		IMAGE Consortium (info@image.lnl.gov) for further information.		
		Insert Length: 1674 Std Error: 0.00		
		Seq primer: Promega -21m13		
		High quality sequence stop: 262.		
FEATURES		Location/Qualifiers		
source		1..452		
		/organism="Homo sapiens"		
		/db_xref="GDB:3770271"		
		/db_xref="taxon:9606"		
		/map="21"		
		/clone="IMAGE:201221"		
		/clone.lib="Soares fetal liver spleen INFLS"		
		/sex="male"		
		/dev_stage="20 week-post conception fetus"		
		/lab_host="DH10B (ampicillin resistant)"		
		/note="Organ: Liver and Spleen; Vector: pTZ19 (Pharmacia)"		

with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer (5' AACTGAGAGATTATTAAGATCTTTTCTTTTCTTTT 3'), double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 106 a 102 c 135 g 102 t 7 others

Query Match 13.7%; Score 393.2; DB 23; Length 452;
Best Local Similarity 96.4%; Pred. No. 4.2e-68;
Matches 432; Conservative 0; Mismatches 12; Indels 4; Gaps 3;

QY 2426 actcttgagccctgggagatcctacatcccgccccccag--tcaagagctcctctgt 2483
DB 452 ACTTGTGAGCCCTGGGAGACCTCAATCCCGTCCCCCAGTMAAAAGCTCTCTGCT 393
QY 2484 cctgccttgatgattcagaggtcagggagtgctcagagctccatctgtatc 2543
DB 392 CTTCCCTTAATGATTCAGGCTCAGGAGTGGCTCAGATCCACTCTCATTTGCTGATC 333
QY 2544 aattctcatcaccctacatctctcagagagccagaccagacagatataccsc 2603
DB 332 AATTCTCTATACCCCTTANACATCTCTCAGAGGACCCAGCCCAAGACATATCCAC 273
QY 2604 aaccccccagccctgagtgtagctgaacctaatgtacgctacatcgtcgttaacag 2663
DB 272 ANACCCCAAGCCCTGAGAGTACTGACCTAATGTGACGCTCATCTCGGTATATCG 213
QY 2664 agatgagcatcccttcacacacagccctt-cccccttctgggtctccatacctagag 2722
DB 212 AGAGTAGCATTCCTTCAATCACAGCCCTTCCCTTCTGGGAGTCTCATATACCTAGAG 153
QY 2723 accactctggaggttggttgtaagccctggagctgcagcgtctgtatggagagatc 2782
DB 152 ACCACTCTGGAGGTGTGCTAGGCCCTGGGACCTGGCAAGCTGTGTAAGGAGATC 93
QY 2783 gctggcaccatagctctatgccaacaggtgctgtgtgtaaggggagtggaattca 2842
DB 92 GCTGGACCATAGCCTTATGGCCAAAGGTGTTGTGTGAAGGGCGTGGAGTTTCA 33
QY 2843 atataataaa-ccacctgatatcaata 2869
DB 32 ATATCATTAACCCACCTGATATCAATA 5

RESULT 7
R46386/c 444 bp mRNA EST 22-MAY-1995
LOCUS y950d07.s1 Soares infant brain INIB Homo sapiens cDNA clone
DEFINITION IMAGE:36272 3', mRNA sequence.
ACCESSION R46386
NID 9822384
VERSION R46386.1 GI:822384
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 444)
Hillier, L., Clark, N., Dubugue, T., Elliston, K., Hawkins, M.,
Holtman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 10, 1995 this sequence version replaced gi:805783.

Contact: Wilson RK

Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 1937
High quality sequence stops: 282 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1937 Std Error: 0.00
Seq primer: Promega -2im13
High quality sequence stop: 282.
Location/Qualifiers

FEATURES
Source

1..444
/organism="Homo sapiens"
/db_xref="GDB:408773"
/db_xref="taxon:9606"
/clone="IMAGE:36272"
/clone_lib="Soares infant brain INIB"
/sex="Female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: lafmid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5'
AATGAGAGATTGCGCGCCGACGAGATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 97 a 92 c 127 g 123 t 5 others

Query Match 13.5%; Score 388; DB 22; Length 444;
Best Local Similarity 97.2%; Pred. No. 4.4e-67;
Matches 413; Conservative 0; Mismatches 10; Indels 2; Gaps 2;

QY 2454 ccccgctcccccagtcacagatcctctctcctcgtcgttaaty-attcagggcagggag 2512
DB 444 CCCCCTCCCCCAATCA-ANCTCCTCTGCTCTGCTGATGATTCAGGGTACGGAG 386
QY 2513 gtgctcagagtcacacctcattgtctgataatttcattaccatcattctcc 2572
DB 385 NTGGCTCAGAGTCCACCTCTCATTCCTGANCATTTCTCATTTACCCCTANACATCTC 326
QY 2573 acggagcccaagcccaagacagatataccacaccccaagccctgagtgatcagacc 2632
DB 325 ACGAGGCCAGACCCACACAGATATCCAAACCCCAAGCCCTGACGTGACGCC 266
QY 2633 taatgtcaggtcactatgctgtgtaatacagagtagcatccctcattacacagccct 2692
DB 265 TAATGTGACGGTCACTGCTGTTAATCAGAGATGATGATTCATTCATCAACACCCCT 206
QY 2693 tccctcttctgggtgccctacatacttagagacctcctggagaggttttctaagccctgg 2752
DB 205 TCCCTTTCTGGGGTCTTCATACCTAGAGACCACTCTGGAGAGTTTCTAGGCCCTGG 146
QY 2753 accgtgcaagctctgttgtagtgagagatcgtgacacatagccttgagcaagagt 2812
DB 145 ACCTGCGCAAGCTCTGTATGTGGAGAGATCCCTGGACACATAGCTTATGGCCAAAGCT 86
QY 2813 gttctgtgtgtaaaaggcgctgaggtttcaatatacaataaaacacactgatacaagc 2872
DB 85 GGTCTGTGTGTAAGAGCGGTGAGATTTCAATATCAATAAACCACTGATATCAATAAA 26
QY 2873 caaaa 2877
DB 25 AAAAA 21

RESULT 8

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1734
High quality sequence stops: 326 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1734 Std Error: 0.00
Seq primer: Promega -21ml13
High quality sequence stop: 326.

Query Match	13.3%	Score 382;	DB 22;	Length 425;
Best Local Similarity	96.7%	Pred. No. 6.7e-66;		
Matches 411; Conservative	0;	Mismatches 11;	Indels 3;	Gaps 2

```

/organism="Homo sapiens"
/db_xref="GDB:1329192"
/db_xref="taxon:9606"
/clone="IMAGE:429423"
/clone_lib="Scares_fetal_liver_spleen_1NFIS_S1"
/sex="male"
/dev_stage="20 week post conception fetus"
/lab_host="H10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subcloned version of the original Scares fetal

```

liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGATTAATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Patricia Bonaldo."

BASE COUNT 74 a 131 c 139 g 89 t 11 others

ORIGIN

Query Match 12.9%; Score 371.8; DB 27; Length 444;
Best Local Similarity 91.9%; Pred. No. 7e-64;
Matches 408; Conservative 0; Mismatches 33; Indels 3; Gaps 2;

620 tggggctgcaatcccgatccggtcac--cccaacaccctgacgtgctgacagagc 677
|||||
444 TGGGCTGCAATTCGGGATCCNGCTCACACACACTGTCAGTGGGTGATGAGGC 385
|||||
678 cgggaagtcggaagagcagctgcgctggaagaccctgactactgccc-tacagcg 736
|||||
384 CGGGAAGTCGGAGAAGAGAGTCCGCTGGAGGACCCTGACGTACTGCTTACAAGC 325
|||||
737 ccatacggaacgtaacggaagagctggtgtacgcccactacgagcgccggaagacctgc 796
|||||
324 NCATGGCAACTCACGGGAGAGCTGTGTACGCCACTACGGGGCCGCGAAGACTGCG 265
|||||
797 aggaacctgagggcagggcgctgcatccagtgccgctgctgtgtgctgcgctgagc 856
|||||
264 AGGACTCGGGGCCAGGGGGCGTGGATCCAGTGGGNCCTGCTGCTGGCGCNGTGGGG 205
|||||
857 tgatcagcttcgcccagaagtgtaacaaatgctcaggaactcggggctaaagaagtgtc 916
|||||
204 TGANTCAGTTCGCCCGAAGGTGACCAATGCTCAGGAGACTTCGGGCTCAAGAGTGTCA 145
|||||
917 tataccagagcagcaggaactctccacagaccaccccaagagcctgttcacagcagc 976
|||||
144 TATACCCAGAGCGAGCGGACTTCTCNCAGAGCCACCCAGCAAGCCGTCTCANNACAGC 85
|||||
977 aggcagcttatgacatgtcaccttggaacttggaacccctacacacctgctccctt 1036
|||||
84 AGGAGTGATGACATNTGCACCTGGGAGACTGAGAGACCCCTACACACTGCTTCCCTT 25
|||||
1037 ccttaacataaaccagttccctc 1060
|||||
24 CCTTCATCAACCCAGNTCCCTC 1

RESULT 10
A1047358 617 bp mRNA EST 08-JUL-1998
LOCUS u655d02.y1 Sugano mouse liver m1a Mus musculus cDNA clone
DEFINITION IMAGE:1450755.5' similar to TR:Q99376 Q99376 TRANSEFERIN RECEPTOR
PROTEIN; mRNA sequence.
ACCESSION A1047358
NID 93295645
VERSION A1047358.1 GI:3295645
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 617)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On May 8, 1995 this sequence version replaced gi:800961.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:924071
Seq primer: custom primer used
High quality sequence stop: 494.
Location/Qualifiers
1. 617
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1450755"
/clone_lib="Sugano mouse liver m1a"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
(CACTGATG); Site_2: DraIII (CAGCATG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCTTACTGG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACTGATG, 3' site CAGCATG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTTTGTCTCTTAAGATGCGC and 3' end
primer CGACCTGCAGCTGACACACA."

BASE COUNT 121 a 177 c 186 g 133 t

ORIGIN

Query Match 12.8%; Score 368.8; DB 41; Length 617;
Best Local Similarity 74.0%; Pred. No. 3e-63;
Matches 534; Conservative 0; Mismatches 82; Indels 106; Gaps 1;

1333 tggggccagagcagctaaatccgctgtgtggagcgtctactcctggagctgtgcg 1392
|||||
1 TGGGCCCCAGGAGCAGCCAACTCTGCACTGGGAGCTGCCAATCTCTGAGACTGTTCGG 60
|||||
1393 accttctcctatggtgagaaagcgttcgagcccgcaagagctcctcttcaagc 1452
|||||
61 ACCTTCTCTTCATGTCAGCAATGGGTTTCAGACTCTGAAAGATCTTTTGTATCAGC 120
|||||
1453 tgggaagctgtgacttgaaagcgttggtcctcagagagtggtcagaagcttaccacg 1512
|||||
121 TGGGAGGAGGTGACTTGGCAGCGGTGGAGCCACAGAGTGGTGGAGGCTTACTCAGC 180
|||||
1513 gtgtcgaactcaaaagccgtatgttaagtgtgagccttggaacaaagcagtgctggagatgac 1572
|||||
181 GTGTCAACACTCAAGAGCTGTGTGACTGAGCTCGA----- 219
|||||
1573 aagttcatgcaagagacagccctcttgcaagctcattgagagtgctcgaagcag 1632
|||||
219 ----- 219
1633 gtgattctcccaacacacagtgggcagactctctatgacagtggtgtcacaatccc 1692
|||||
219 -----CAGACCTCTATGAAAGTGGGACCTCACCACCCC 254
|||||
1693 agctggagatgctgagtgatccggccctaccatggaagcagtgctactatccctcaag 1752
|||||
255 AGCTGGGATGCTGAAGTATCAGCCCTCGCCATGGACACAGATGATATTCCTTCA 314
|||||
1753 gctttggagagtcctcgctgagatctcctcttcttgagagcagcagcctaccaca 1812
|||||

Db 315 GCCTTTCGGGGGTCACAGTGTGGAGTTCCTTCATGAGATGATCGGCTGACCA 374
 QY 1813 ttctgtcacacaaaggaggaacttaagagaacctgcaagtgtcgaagccgctg 1872
 Db 375 TTCTGTGACACAGGAGGAGACATATGAGATCTGCACAGATGCTCGAGTCCGCTG 434
 QY 1873 cccgcctgtgcccagggcctgtccagctgcccagggccctccatccgctcagccac 1932
 Db 435 CCCCCCTGTGTCACAGGAGTGTGCTCAGTCCGGGCCACGCTCTCATCCAGTACGCCAC 494
 QY 1933 gatcgctgtgcccctcagactcgcgcgtacgagagcgtcgtcctcaagcacatcg 1992
 Db 495 GATCACTACTACGCGCTAGACTTGCGCGCATATGAGACGTGGTTCACGACATCGGC 554
 QY 1993 aacctcagaagttctctgtggagacctaaagcccggtgacctgcaagtgtgtac 2052
 Db 555 AACCTCATAGTTCCTGTGCGGACCTCAAGGAGCGCGGCTGACCTGCAATGAGTGAC 614
 QY 2053 tc 2054
 Db 615 TC 616

RESULT 11
 AA505641/c 384 bp mRNA EST 18-AUG-1997
 LOCUS nh82403.s1 NCI_CGAP_Br1.1 Homo sapiens cDNA clone IMAGE:964972 3'
 DEFINITION mRNA sequence.
 ACCESSION AA505641
 NID 92241778
 VERSION AA505641.1 GI:2241778
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 384)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 9, 1995 this sequence version replaced gi:802599.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www.bio.llnl.gov/bdrrp/image/image.html

Insert Length: 2525 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 256.

FEATURES

Location/Qualifiers
 1..384
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:964972"
 /clone_id="NCI_CGAP_Br1.1"
 /sex="female, pooled"
 /tissue.type="breast"
 /lab_host="DH10B"
 /note="Vector: pT773D-Pac (Pharmacia) with a modified
 polylinker: 1st strand cDNA was prepared from pooled bulk
 breast tumor tissue, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT773

BASE COUNT 85 a 86 c 108 g 105 t
 ORIGIN
 vector. Library is not normalized. (The normalized
 version of this library is NCI_CGAP_Br2.) Library was
 constructed by Bento Soares and M. Fatima Bonaldo."

Query Match 12.6%; Score 361.4; DB 34; Length 384;
 Best Local Similarity 98.4%; Pred. No. 7.5e-62;
 Matches 365; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2507 agggaggtgtgctcagagctcacctctcatgtcgtatcaattctcattaccctacacat 2566
 Db 384 AGGAGGTGTGCTCAGATTCACCTCTCATGCTGATCAATTTTCATTACCCCTACACAT 325
 QY 2567 ctctccagagagcccaagcccaagacagatattccacaccccaagccctgcaagttagc 2626
 Db 324 CTCGCCAGGAGGCCACACCCACGACGATATCCACACACCCAGCCCTGCAAGTATAGC 265
 QY 2627 tgaccctaattgtacggttactgtcgttgaatcagagagtagcatccctcaatcaca 2686
 Db 264 TGACCCCTAATGTGACGGGTCTACTGTGCTTAATCAAGAGATGACATCCCTCAATCACA 205
 QY 2687 gcccttcccttcttctgtggtctctccattcagagacccactctggaggttgcctaacg 2746
 Db 204 GCCCTTCCCTTCTCTGTGGGTCTCTCCATACCTAGAGACCACTCTGGAGGTTTGTGCTAGGC 145
 QY 2747 ccttgagacctgtgccaagctcgttagtgtagagagatcctgtgcccataagcttagtgcga 2806
 Db 144 CCTGGAGACTGTCGCGCAGCTCTGTATGTGGAGATGCTGTGGACCATAGCCTTATGGCCA 85
 QY 2807 acagggtgtctgtgtgtaaaagggtggtgagttcaataatcaataaaccactgatalca 2866
 Db 84 ACAGGTGCTGTGTGTGAAGGGCGCTGAGTTTCATATCATCAATAAACACCTGATATCA 25
 QY 2867 ataagcccaaaa 2877
 Db 24 ATAAAAA 14

RESULT 12
 AA682374/c 388 bp mRNA EST 19-DEC-1997
 LOCUS zj86a08.s1 Soares_fetal_liver_spleen_INFIS_S1 Homo sapiens cDNA
 DEFINITION clone IMAGE:461750 3' similar to gb:M11507 TRANSFERRIN RECEPTOR
 PROTEIN (HUMAN);, mRNA sequence.

ACCESSION AA682374
 NID 92686655
 VERSION AA682374.1 GI:2669655
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 388)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project
 Unpublished (1997)
 JOURNAL On Sep 12, 1996 this sequence version replaced gi:1394174.
 COMMENT

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 380.

Db 63 GAAAGGGCGGTGAGTTTCATATCAATAAACCCGATCATATAAAAAA 9

RESULT 14
LOCUS H63101/c 384 bp mRNA EST 11-OCT-1995
DEFINITION YR50H04.s1 Soares fetal liver spleen INF1S Homo sapiens cDNA clone
(HUMAN);, mRNA sequence.

ACCESSION H63101
NID 91017902
VERSION H63101.1 GI:1017902
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 384)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Matra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
COMMENT On May 8, 1995 this sequence version replaced gi:799827.

TITLE
JOURNAL
COMMENT

FEATURES
source
1.384
/organism="Homo sapiens"
/db_xref="GDB:3177890"
/db_xref="taxon:9606"
/clone="IMAGE:208759"
/clone_lib="Soares fetal liver spleen INF1S"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pRT73D (Pharmacia)
with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Patrina Bonaldo."

BASE COUNT 66 a 104 c 126 g 78 t 10 others

ORIGIN

Query Match 11.6%; Score 334.6; DB 24; Length 384;
Best local Similarity 94.5%; Pred. No. 1.4e-56;
Matches 362; Conservative 0; Mismatches 19; Indels 2; Gaps 2;

QY 711 ccctgacgtctactgcccctacagcgcacatcggaacgtacggaagctgtgtcgcg 770
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 383 CCCTGANGTCTACTGNCCTAAAGCGCNCATCGMAACGTCA-GGAGAGNCCTGGGTAGC 325
QY 771 ccactacggcgccgcaagaactgcagacctgcggcgccagggcgctgtgattccacgtgg 830

Db 324 CNACTACGGGNGGCGCCGAACACCTGCAGACCTCGCGGCCAGGGGCTGAGATCACTGGG 265
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
QY 831 ccgctcgtctgtgtgctgcgtggtgggtg-atcagcttcgccagaagtgcacatgctc 889
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 264 CCGCTGTGCTGTGTGTGCGGGGTGAGTAATACANNTCCGCCAGAGGTGACCAATGCTC 205
QY 890 aggaactcggggcctcaaggaagtgtcatataccagaagccagcgaactctccagaacc 949
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 204 AGGACTTCGGGGCTCAAGAGAGTCTCATATACCAAGAGCCAGNGACTTCTCCAGAGCC 145
QY 950 caccacaagccaagctcgttcagcagcagcagcagtgatgagatgacatgacatgagac 1009
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 144 CACCCAAAGCCAGACCTGTCCAGCAGCAGCATGTATGAGACTGTGCNCTGGAACTG 85
QY 1010 gaagaccctacacacactgcttccttcctcaatcaaacccagctccctccagtgcat 1069
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
Db 84 GAGACCCCTACACACCTGGCTTCCTTCTCATATCAAAACCCAGTTCCTCCAGTGCAT 25
QY 1070 catcagggcttccagcatccca 1092
Db 24 CATCAGGCGCTTCCACACATCCCA 2
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||

RESULT 15
LOCUS H70432/c 366 bp mRNA EST 24-OCT-1995
DEFINITION YS03C12.s1 Soares fetal liver spleen INF1S Homo sapiens cDNA clone
IMAGE:213718 3', mRNA sequence.

ACCESSION H70432
NID 91040638
VERSION H70432.1 GI:1040638
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 366)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapellil,B.,
Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,M.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellendberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Matra,M.
Generation and analysis of 280,000 human expressed sequence tags
97044478
On Apr 14, 1993 this sequence version replaced gi:837502.

TITLE
JOURNAL
MEDLINE
COMMENT

FEATURES
source
1.366
/organism="Homo sapiens"
/db_xref="GDB:3179022"
/db_xref="taxon:9606"
/map="11"
/clone="IMAGE:213718"
/clone_lib="Soares fetal liver spleen INF1S"
/sex="male"

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Insert Size: 801
High quality sequence stops: 269
Source: IMAGE Consortium, LNL.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 801 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 269.
Location/Qualifiers

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 13, 2000, 02:17:22 : Search time 1880.46 Seconds

(without alignments)
6541.900 Million cell updates/sec

Title: US-09-358-755-2

Perfect score: 2877

Sequence: 1 ctgcagcgtctcagagagga.....ctgatatacaataagccaaaa 2877

Scoring table: IDENTITY_NUC

Searched: 780561 seqs, 2137953050 residues

Database: GenEmbl.*

Word size: 0

Number of hits that pass the threshold : 1561122

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: gb_v1:*
17: em_fun:*
18: em_htg:*
19: em_hum1:*
20: em_hum2:*
21: em_in:*
22: em_om:*
23: em_or:*
24: em_ov:*
25: em_pat:*
26: em_ph:*
27: em_pl:*
28: em_ro:*
29: em_sts:*
30: em_sy:*
31: em_un:*
32: em_v1:*
33: gb_htg1:*
34: gb_htg2:*
35: gb_in1:*
36: gb_in2:*
37: em_ba1:*
38: em_ba2:*
39: em_hum3:*
40: em_hum4:*
41: gb_pr4:*
42: gb_htg3:*
43: gb_htg4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2875.4	99.9	2877	41	AF067864	AF067864 Homo sapi
2	711.8	24.7	227968	11	AF053356	AF053356 Homo sapi
3	550.4	19.1	580	11	HUMR50H04	AF085928 Homo sapi
4	305.4	10.6	2500	4	GCTER	X55348 G.gallus mr
5	280.8	9.8	5010	9	HSTR	X01060 Human mRNA
6	280.2	9.7	2826	10	HUMTFRR	M11507 Human trans
7	256	8.9	2460	12	CRUTRANREC	L19142 Cricetus
8	247	8.6	3413	12	RATFRFR	M58040 Rat transfe
9	215	7.5	2292	12	MMTRMRNA	X57349 M.musculus
10	160.2	5.6	780	5	I23809	I23809 Sequence 27
11	146	5.1	250	13	G14981	G14981 human STS S
12	127	4.4	540	5	I23811	I23811 Sequence 29
13	103.6	3.6	2570	12	AF009921	AF009921 Rattus no
14	91.6	3.2	2320	10	HSAN012371	AJ012371 Homo sapi
15	88.8	3.1	660	5	I23810	I23810 Sequence 28
16	73.4	2.6	3152	10	HSAN012370	AJ012370 Homo sapi
17	67.8	2.4	1428	12	AF039707	AF039707 Rattus no
18	67.8	2.4	2348	12	AF040256	AF040256 Rattus no
19	67.8	2.4	2899	12	RNU75973	U75973 Rattus norv
20	66.8	2.3	2603	12	AF026380	AF026380 Mus muscu
21	57.8	2.0	1278	1	ABIPDAGEN	X88853 A.brasiliens
22	56.6	2.0	43632	1	SC6A5	AL049485 Streptomy
23	56.2	2.0	2532	3	AF050502	AF050502 Sus scrof
24	56	1.9	22449	1	SPSNBCDE	Y11548 S.pistillae
25	56	1.9	22449	1	SPSNBCGEN	X98690 S.pistillae
26	55.6	1.9	306	13	G45215	G45215 221292.1 ze
27	54.8	1.9	189	12	S81327	S81327 transferlin
28	53.8	1.9	2653	5	I23794	I23794 Sequence 1
29	53.8	1.9	2653	10	HUMPSM	M99487 Human prost
30	53.8	1.9	2518	41	AF176574	AF176574 Homo sapi
31	53.6	1.9	4459	1	TFPHST	X65609 T.thermophi
32	53.6	1.9	7621	1	TFPHSTRO	Y15464 Thermus the
33	53.6	1.9	4171	1	TFPHSTG	Z12118 T.thermophi
34	53	1.8	38995	1	SCF34	AL109974 Streptomy
35	53	1.8	1189	9	GCTRBPX1	Z11805 G.circalicau
36	51.8	1.8	3508	2	AF069761	AF069761 Halobacte
37	51.8	1.8	2014	2	AZSID	I26240 Azospirillum
38	51.8	1.8	982	7	ZMHSP18K1	X54075 Maize mRNA
39	50.2	1.7	10295	1	D63799	D63799 Thermus the
40	50.2	1.7	40745	1	SC151	AL109848 Streptomy
41	49.8	1.7	24700	1	SCB8	AL035654 Streptomy
42	48.8	1.7	40105	1	SC5F2A	AL049587 Streptomy
43	48.8	1.7	26440	1	SC15	AL049707 Streptomy
44	48.6	1.7	843	7	ZMHSP18K2	X54076 Maize mRNA
45	48.2	1.7	4933	1	ABIPDC	X9587 A.brasiliens

ALIGNMENTS

RESULT	1	LOCUS	AF067864	2877 bp	mrna	PRI	27-JUL-1999
DEFINITION							Homo sapiens transferrin receptor 2 alpha (TFR2) mRNA, complete cds.
ACCESSION			AF067864				
NTID			g5596369				
VERSION			AF067864.1	GI:5596369			
KEYWORDS							
SOURCE							human.
ORGANISM							Homo sapiens
REFERENCE							Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS							1 (bases 1 to 2877) Kawabata,H., Yang,R., Hirama,T., Vuong,P.T., Kawano,S., Combarat,A.F. and Koefler,H.P.
TITLE							Molecular cloning of transferrin receptor 2. A new member of the transferrin receptor-like family
JOURNAL							J. Biol. Chem. 274 (30), 20826-20832 (1999)


```

Db 1501 GGGTACTGACGCGTGGCGACCTCAAGCCGTAGTGTAGCTGAGCCCTGGACACAGCGACTG 1560
Qy 1561 ctggggagatyaagaatttcacaaagacagccccccttcgaagcttcattagagt 1620
Db 1561 CTGGGGAGATGACAAAGTTTCATGCCAAGACAGCCCTTCTGACAAAGTCTCATGTAGAGT 1620
Qy 1621 gtcttgaaagagtgatgtttcccaacaagatgggagagctctctatgaacagtggtg 1680
Db 1621 GTCTTGAAGAGGTGATTTCTCCAAACACAGTGGGAGACTCTATGTAAGAGTGTGTG 1680
Qy 1681 ttcaacaatcccaagctggagatgctgtaagtgatccgagccctacacatggagacaagtgcc 1740
Db 1681 TTACCAATCCCAAGTGGAGTGTGAGAGTATCCGGCCCTTACCAATGAGACAGAGTGTG 1740
Qy 1741 taattcctcaagagcttggtggagtgcttcgctgcagtgagttcttccttatggagagac 1800
Db 1741 TATTCTTCAAGCGCTTTGTGGAGTCCCTGCGTCCGAGTTCTCTTATGAGAGACGAC 1800
Qy 1801 cagcctaccatcttcctgcacacaagaagagacattatagaacctgataagtgctg 1860
Db 1801 CAGGCTTACCCATTCCTGCAACAAAGAGAGACATTATGAAACCTGCAATAGTGTCTG 1860
Qy 1861 caagagccctgcgcgcgcgtggcccaagcgctggccagcttcgaaaggagagcttcctatc 1920
Db 1861 CAAGCGCGCTGCGCGCGCGTGGCCAGGCGCTGCGCCAGCTCGAGGAGCGACTCTCTATC 1920
Qy 1921 cggctcaagcagatcgctgctgccccctgacttcgcccgtacagggagagctgctc 1980
Db 1921 CGGCTCAGCCAGATCGCTGCTGCCCCCTGACTTCCGCGCTACGGGAGAGTGTCTCTC 1980
Qy 1981 aggcacatcggaagacccaacagagttctcgggagactcaaggcccgagctgcagctg 2040
Db 1981 AGGCACATCGGGAACCTCAAGAGTTCTCGGGAGACTCAAGGCCCGCGGCTGACCTG 2040
Qy 2041 cagtgagtgtaactcgcgcgaggagactatccggcgagcgaaagagctgcgcagagag 2100
Db 2041 CAGTGGGTGTACTCGCGCGGGGGGAGTACTCCGGCGGCGGAAAGCTGCGGACAGAG 2100
Qy 2101 atctcaagctcgagagagagagacagagcagctgacacagctgtaacaagctgcgata 2160
Db 2101 ATCTTCAAGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Qy 2161 cgggtgagtgtaacttccttcctccagtaagtgctgcagagcagagctcccgctccgac 2220
Db 2161 CGGGTGAGTGTACTTCTTCCAGTACGTGCGCACCGACCGACTCCCGTCCGCGCAC 2220
Qy 2221 atcttcagcgctggagagacacagctggcgccctgctgagacacacctgcgagctg 2280
Db 2221 ATCTTATGGGCGCTGGAGACACACACCTGGGCGCTGCTGGACACCTGGGCTGCTG 2280
Qy 2281 cgcctcaacagctcgcggagaccccgaggagacactctctcactggtcttcagagagcgt 2340
Db 2281 CGCTCAACAGCTTCGGGAGACCCCGGGGACACCTCTCACTGGCTTCAAGAGAGCGT 2340
Qy 2341 ttccgagctagctagcctgcctcactgagcgtcgaagggagcgaatgagctaac 2400
Db 2341 TTCCGGGCTAGCTAGCCCTGCTCACCTGACGCTGCAAGGGGAGCAATGCTTATGC 2400
Qy 2401 gggagatgtcggagacattgataaactttgagggcctgggagttcttcacatcccgctc 2460
Db 2401 GGGAGATGTCTGGAAATTTGATTAACAACTTCTGAGGCGCTTGGGATCTCTCAATCCCG 2460
Qy 2461 ccccgagcagaagctcctcctgctcctcgtctgtaagttcagggctcagggagtgctca 2520
Db 2461 CCCAGTCAAGAGCTCTCTGCTCTGCTCTGATGATGATTCAGGGTCAAGGGAGTGCTCA 2520
Qy 2521 gagttcaacctcatgtgctatcaatttctcaataccctacacatctctccagagcc 2580
Db 2521 GAGTTCACCTCTATTCCTATTCATTAATTCATTAATTCATTAATTCATTCATTCATTCAT 2580
Qy 2581 cagagccagcagagatcatcacacacccagccctgagtgtagctgacccataatgta 2640

```

```

Db 2581 CAGACCCAGCAGATATCCACACACACCCAGCCCTGTCATGTAGCTGAGCCCTATATGTA 2640
Qy 2641 cggctacactgctggttaatcagagagtagatcccttcacatcagagcccttcctc 2700
Db 2641 CGGCTATACTGTCGGTTAATCAGAGAGTAGATCAATCCCTTCAATCAACAGCCCTTCCCTTT 2700
Qy 2701 ctgggttccctcaactactaagagacacactcgtggaggttttctaaagccttgagctggc 2760
Db 2701 CTGGGTCCTTCATTAATCAAGACACTCTGGAGGTTTCTTGGCCCTGGAGCTGGCC 2760
Qy 2761 agctctgtagtgagagatcgcctgcacataagcctctatggccaacagtgagctg 2820
Db 2761 AGCTCTGTATGTGGAGAGATCCCTGCGACACATAGCCTTATGGCCAAAGTGTGTG 2820
Qy 2821 gtgaagggcggtggagtttcaatataaataaacacactgatatcaataagccaana 2877
Db 2821 GTGAAGGGCGGTGGAGATTTCATATCAATAAACACCTGATATCAATAGCCAAA 2877

RESULT 2
AF053356 227968 bp DNA PRI 24-NOV-1998
LOCUS Homo sapiens chromosome 7q22 sequence, complete sequence.
DEFINITION AF053356
ACCESSION AF053356
NID G3135305
VERSION AF053356.1 GI:3135305
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 227968)
AUTHORS Gloeckner,G., Scherer,S., Schattevoy,R., Boright,A., Weber,J.,
Tsui,L.C. and Rosenthal,A.
TITLE Large-scale sequencing of two regions in human chromosome 7q22:
analysis of 650 kb of genomic sequence around the EPO and CTR1
loci reveals 17 genes
JOURNAL Genome Res. 8 (10), 1060-1073 (1998)
MEDLINE 99018118
REFERENCE 2 (bases 1 to 227968)
AUTHORS Gloeckner,G., Rosenthal,A., Drescher,B., Weber,J. and Schattevoy,R.
TITLE Direct Substitution
JOURNAL Submitted (05-MAY-1998) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q22"
/region="Region: 3763"
/feature="SINE/Alu, pos., in, rep.: 2. .197"
/rpt_family="Alu"
1.169722
/feature="PAC clone 3763"
/feature="Region: 3763"
complement(join<309. .578,2314. .2458,3620. .3831,
14784. .14928,15156. .15357,16624. .16804,18675. .18769,
19928. .20039,20325. .20562,21131. .21378,21722. .21863,
22834. .23036,24183. .24361,24472. .24582,26030. .26173,
27852. .27997,28543. .28667,29738. .29857,30045. .30186,
32978. .33158,35152. .36676,37121. .37205,37491. .37762,
41053. .41140,41322. .41593,41773. .41919,46462. .46608,
46717. .46795,47291. .47382,48230. .>48478)
/product="zonadhesin"
complement(309. .378)
/feature="100.000, excellent"
complement(join<309. .578,2314. .2458,3620. .3831,
8786. .9180,12095. .12222,12821. .12985,13079. .13326,
14784. .14928,15156. .15357,16624. .16804,18675. .18769,
19928. .20039,20325. .20562,21131. .21378,21722. .21863,
22834. .23036,24183. .24361,24472. .24582,26030. .26173,

```

```

27852..27997,28543..28667,29738..29857,30045..30186,
32978..33158,35152..36676,37121..37205,37491..37762,
41053..41140,41382..41593,41773..41913,44662..44608,
46717..46795,47291..47382,48230..48383))
/codon_start=1
/product="zonadhesin"
/protein_id="AAC78790.1"
/db_xref="PID:g3135306"
/db_xref="GI:3135306"
/translacion="MLDLPKNARPGQKAVLLSPVSLSSGCLSFSEFHYILRQSPGAA
HIVASVLSIRKHTLFGSDPGPNQAVSVNTAVRQIQFAVVGFKPEEAVAVDT
SIAPGEGFPQCCEDENAHPCDMVOTSGDGHGNAHKGKNGPVHGAGGAFPNADL
TQCPEDDAPKLCDSMOYSADDEDMVRASGSPGSHGACPGYNGEGSYHMSNSP
HRGSVARLSPDLMEQGPLCYHFAHMGKLSWGQOLRLLLSGEGRRPDYLMKHMNT
ORPSMMLTVTPAGFPLPLRLMEFGSTGSAVYLDIALDALSIKRGSCNRGHA1YLE
ADESHAGQSVRLVSPGAGDLCVERAVAYMGEGTMTLLIGSPAGSPPIPLMK
RVSQRPWQNTSVTVSGHQPMOLIFKGIQGSNTASVAMGFLINPGCPKAVLP
ELPVPVSTSGPSETTGLTENPITSTKKPTVSIKPKPTPEKPTPEKPTPEKPTPEK
TISTEKPTPEKPMSEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEK
PTEPTPTPEETTYMEBPVPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEK
EKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEK
LTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEK
TSVEETITSEKLTIPMEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
IEETITSEKLTIPTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
KPTISPEKLTIPTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
PSPATGALVMSRPAHSPMTSPVITCTTSSSTGMSCPNARYESCPCASCKS
PRSPGCLRCREGCCVNFELSDNHCIOASSCCYNDIYEPQAKFETPCTHEKRC
WPSRVEQIISQCGTHVCQLKNGOYGHYPAGATLVYDHPYVFDGHHFPMK
CTYILAPKSGVSTDFPVYAKNEQGEQVSCISKAVYLPBESTVTLKLRRLVVG
QOVLPIAPKGVFLGASGREVLQTEFGLRVDQDLVYVTSYSGCLCLGN
YDGSNDNHLKLDSPAGDKDELGNMQTDDEQECQKQVNVSPSCDSSLQSGMS
PGFCGLVDTHGPFETCLLHYKASFPDSCMLDMCGROGLHLLCTHMTSTTTCOA
GHAYKPMREPFCELELSMACPPNRSKYSICAPPCDTCGSGFSCFSCRYEACECP
EFGVSGECLIPRSQGLHPAGSFTFKGERMYKQCKELCVCSNNRIRPCQWECRQD
EFCQOQD6IYCHQAGATCTASGDPHVLTEDGALHHRMGCTVYLRPCQWECRQD
FVVSATNENRGILFVSYIKAVHTVDELISILRGCVMKNGHVALPVYLAQGRV
IRLSNVLTLTYNFGQYKDYSHLVELTVPSYGGOLCGEFGHLLRKAQVMT
RVTVNNNSLDDNLRPDKRLAGDSMOGAAMKLPESSEGCFLVYGGKSPSCONSNA
DANKNCAILLNPGPSQCHQVYVPOSSPASCYHGGCGTGTATLACRSIQATASLC
AQQAQAPAMRNKRTFCRECPGSSVSPSCSPDTCSSINNRPDCKPALPCASESCOK
GHILSGTSCVPLHCRASGVQCLPGESHVYSPDSNHSIPACTLVLYKCHPMAALP
NOICWALDGLHCRASGVQCLPGESHVYSPDSNHSIPACTLVLYKCHPMAALP
LEKISAKHEEGEGTEAFRLHYIDMVAQVTLPTQYKGRVYLNKQVTLPIISQIPG
SVKSSSYITVNIKIGOVKREHYIDMVAQVTLPTQYKGRVYLNKQVTLPIISQIPG
DEKLANDSEFVNSMKDKDIPSCOSLPVDQOQIPABQENPSCGRADLRAREK
EAALRAPVWQACASRIDLTPFLVCANTLCFEGLYQALCALQAFATQCSQSLKRP
LMRNSFCHECALLSPQPPKPPPLCFPESLTLPLVPLECPAYSSYNTNLPSCSPS
CWDLDGRCGAKVPSACAEQICQPGYVLSDEKCVPRQCCCKDAHGGSYPLGKSW
SGCTEKCVCTGAIQCGDFRCPSGHQCLTSDNSNSNCVSKSQCVCYGPRLTLD
GFSYRLOGRMVYLIKTYDVLPEGEVPLVGRKMKDPRRSITLQEVITTYVYKYO
LQAGLELYVRAGPGP"
312..335
exon
/note="89..000, shadow-excellent"
/complement(327..435)
/misc_feature
/note="99, 1:xyz, 1..109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, length, =,
1090, P, =, 2.3e-37, S, =, 99"
/note="Region: Data base match"
/complement(327..435)
/misc_feature
/note="99, 173:xyz, 1..109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, length, =,
1090, P, =, 1.1e-32, S, =, 99"
/note="Region: Data base match"
/complement(327..435)
/misc_feature
/note="99, 1:xyz, 1..109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, length, =,
1090, P, =, 2.3e-37, S, =, 99"
/note="Region: Data base match"
/complement(328..435)
/misc_feature
/note="100, 3614:xyz, 1..106, of, GB|U83191|HSU83191_1,
Homo, sapiens, Eukaryotae, mitochondrial, eukaryotes,
Metazoa, Human, zonadhesin, (ZAN), mRNA, partial, cds.,
Sperm, membrane, protein., 4/97, length, =, 363, P, =,
27852..27997,28543..28667,29738..29857,30045..30186,
32978..33158,35152..36676,37121..37205,37491..37762,
41053..41140,41382..41593,41773..41913,44662..44608,
46717..46795,47291..47382,48230..48383))
/codon_start=1
/product="zonadhesin"
/protein_id="AAC78790.1"
/db_xref="PID:g3135306"
/db_xref="GI:3135306"
/translacion="MLDLPKNARPGQKAVLLSPVSLSSGCLSFSEFHYILRQSPGAA
HIVASVLSIRKHTLFGSDPGPNQAVSVNTAVRQIQFAVVGFKPEEAVAVDT
SIAPGEGFPQCCEDENAHPCDMVOTSGDGHGNAHKGKNGPVHGAGGAFPNADL
TQCPEDDAPKLCDSMOYSADDEDMVRASGSPGSHGACPGYNGEGSYHMSNSP
HRGSVARLSPDLMEQGPLCYHFAHMGKLSWGQOLRLLLSGEGRRPDYLMKHMNT
ORPSMMLTVTPAGFPLPLRLMEFGSTGSAVYLDIALDALSIKRGSCNRGHA1YLE
ADESHAGQSVRLVSPGAGDLCVERAVAYMGEGTMTLLIGSPAGSPPIPLMK
RVSQRPWQNTSVTVSGHQPMOLIFKGIQGSNTASVAMGFLINPGCPKAVLP
ELPVPVSTSGPSETTGLTENPITSTKKPTVSIKPKPTPEKPTPEKPTPEKPTPEK
TISTEKPTPEKPMSEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEK
PTEPTPTPEETTYMEBPVPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEK
EKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEK
LTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEKPTPEK
TSVEETITSEKLTIPMEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
IEETITSEKLTIPTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
KPTISPEKLTIPTEKPTISTEKPTISTEKPTISTEKPTISTEKPTISTEKPTIP
PSPATGALVMSRPAHSPMTSPVITCTTSSSTGMSCPNARYESCPCASCKS
PRSPGCLRCREGCCVNFELSDNHCIOASSCCYNDIYEPQAKFETPCTHEKRC
WPSRVEQIISQCGTHVCQLKNGOYGHYPAGATLVYDHPYVFDGHHFPMK
CTYILAPKSGVSTDFPVYAKNEQGEQVSCISKAVYLPBESTVTLKLRRLVVG
QOVLPIAPKGVFLGASGREVLQTEFGLRVDQDLVYVTSYSGCLCLGN
YDGSNDNHLKLDSPAGDKDELGNMQTDDEQECQKQVNVSPSCDSSLQSGMS
PGFCGLVDTHGPFETCLLHYKASFPDSCMLDMCGROGLHLLCTHMTSTTTCOA
GHAYKPMREPFCELELSMACPPNRSKYSICAPPCDTCGSGFSCFSCRYEACECP
EFGVSGECLIPRSQGLHPAGSFTFKGERMYKQCKELCVCSNNRIRPCQWECRQD
EFCQOQD6IYCHQAGATCTASGDPHVLTEDGALHHRMGCTVYLRPCQWECRQD
FVVSATNENRGILFVSYIKAVHTVDELISILRGCVMKNGHVALPVYLAQGRV
IRLSNVLTLTYNFGQYKDYSHLVELTVPSYGGOLCGEFGHLLRKAQVMT
RVTVNNNSLDDNLRPDKRLAGDSMOGAAMKLPESSEGCFLVYGGKSPSCONSNA
DANKNCAILLNPGPSQCHQVYVPOSSPASCYHGGCGTGTATLACRSIQATASLC
AQQAQAPAMRNKRTFCRECPGSSVSPSCSPDTCSSINNRPDCKPALPCASESCOK
GHILSGTSCVPLHCRASGVQCLPGESHVYSPDSNHSIPACTLVLYKCHPMAALP
NOICWALDGLHCRASGVQCLPGESHVYSPDSNHSIPACTLVLYKCHPMAALP
LEKISAKHEEGEGTEAFRLHYIDMVAQVTLPTQYKGRVYLNKQVTLPIISQIPG
SVKSSSYITVNIKIGOVKREHYIDMVAQVTLPTQYKGRVYLNKQVTLPIISQIPG
DEKLANDSEFVNSMKDKDIPSCOSLPVDQOQIPABQENPSCGRADLRAREK
EAALRAPVWQACASRIDLTPFLVCANTLCFEGLYQALCALQAFATQCSQSLKRP
LMRNSFCHECALLSPQPPKPPPLCFPESLTLPLVPLECPAYSSYNTNLPSCSPS
CWDLDGRCGAKVPSACAEQICQPGYVLSDEKCVPRQCCCKDAHGGSYPLGKSW
SGCTEKCVCTGAIQCGDFRCPSGHQCLTSDNSNSNCVSKSQCVCYGPRLTLD
GFSYRLOGRMVYLIKTYDVLPEGEVPLVGRKMKDPRRSITLQEVITTYVYKYO
LQAGLELYVRAGPGP"
312..335
exon
/note="89..000, shadow-excellent"
/complement(327..435)
/misc_feature
/note="99, 1:xyz, 1..109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, length, =,
1090, P, =, 2.3e-37, S, =, 99"
/note="Region: Data base match"
/complement(327..435)
/misc_feature
/note="99, 173:xyz, 1..109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, length, =,
1090, P, =, 1.1e-32, S, =, 99"
/note="Region: Data base match"
/complement(327..435)
/misc_feature
/note="99, 1:xyz, 1..109, of, GB|U83191|HSU83191, Human,
zonadhesin, (ZAN), mRNA, partial, cds., 4/97, length, =,
1090, P, =, 2.3e-37, S, =, 99"
/note="Region: Data base match"
/complement(328..435)
/misc_feature
/note="100, 3614:xyz, 1..106, of, GB|U83191|HSU83191_1,
Homo, sapiens, Eukaryotae, mitochondrial, eukaryotes,
Metazoa, Human, zonadhesin, (ZAN), mRNA, partial, cds.,
Sperm, membrane, protein., 4/97, length, =, 363, P, =,
1.9e-14, S, =, 100"
/note="Region: Data base match"
/complement(328..576)
/misc_feature
/note="81, 6:xyz, 5899..6145, of, GB|U40024|SSU40024_1,
Sus, scrofa, Eukaryotae, mitochondrial, eukaryotes,
Metazoa, Sus, scrofa, zonadhesin, (Zan), mRNA, complete,
cds., The, amino, acid, sequence, predicts, seven,
extracellular, domains., Including, a, mucin-like, domain,
and, five, domains, homologous, to, the, D-do., ' '
11/95, length, =, 2476, P, =, 0.0e+00, S, =, 81"
/note="Region: Data base match"
/complement(331..578)
/misc_feature
/note="24..00, 82aa,
SEQSVTGDPRYLTFDFGSTRLOGRMVYLIKTYDVLPEGEVPLVGRKMKDPRRS
IFLQEVITTYVYKVOQLQALELY"
complement(331..578)
/note="1..000, 0.62, 0.44, 0.40, 1.22, 0.54, 0.57, 0.60"
complement(339..588)
/promoter
/note="7.43, TATA:, 367, CAP:, 337"
complement(345..561)
/note="98, prediction:, 0%, 98%"
complement(403..570)
/note="80, 3119:xyz, 1..166, of, GB|U83190|MMU83190_1,
Mus, musculus, Eukaryotae, mitochondrial, eukaryotes,
Metazoa, Mus, musculus, zonadhesin, (Zan), mRNA,
partial, cds., Sperm, membrane, protein., 4/97, length, =,
513, P, =, 2.0e-24, S, =, 80"
/note="Region: Data base match"
complement(475..567)
/note="70, 3642:xyz, 856..946, of, GB|U94132|HUMIMCB_1,
Homo, sapiens, Eukaryotae, mitochondrial, eukaryotes,
Metazoa, Human, mucin, 2, (MUC2), mRNA, sequence., 1/95,
length, =, 984, P, =, 3.2e-14, S, =, 70"
/note="Region: Data base match"
complement(475..567)
/note="70, 1807:xyz, 13441..13531, of,
GB|U121998|HMMUC2X_1, Homo, sapiens, Mammalia, Homo,
Metazoa, Chordata, Vertebrata, Mammalia, Homo,
sapiens, Intestinal, mucin, (MUC2), mRNA, complete, cds.,
1/98, length, =, 5179, P, =, 1.5e-30, S, =, 70"
/note="Region: Data base match"
606..903
/note="SINE/Alu, pos., in, rep.: 1..296"
/rpt_family="Alusx"
1004..1039
/note="83..30, %, 22, length:, 2, #: 18, seq:, tt"
/rpt_type=tandem
/rpt_unit=score:
complement(1021..1133)
/note="SINE/Alu, pos., in, rep.: 3..300"
/rpt_family="Aluuo"
complement(1374..1684)
/note="SINE/Alu, pos., in, rep.: 1..302"
/rpt_family="Aluuo"
1745..2050
/note="SINE/Alu, pos., in, rep.: 1..302"
/rpt_family="Aluuo"
2026..2275
/note="7.13, TATA:, -1, CAP:, -1"
2166..2259
/note="14..60, 354, bp"
/promoter
/note="Region: GC Content"
complement(2172..2421)
/note="5..62, TATA:, -1, CAP:, -1"
complement(2314..2458)
/note="0..850, 0.45, 0.46, 0.56, 1.11, 0.49, 0.54, 0.53"
complement(2314..2458)
/note="94..000, excellent"
complement(2354..2464)
/note="75, 26:xyz, 5749..5857, of, GB|U40024|SSU40024_1,
Sus, scrofa, Eukaryotae, mitochondrial, eukaryotes,
Metazoa, Sus, scrofa, zonadhesin, (Zan), mRNA, complete,
cds., The, amino, acid, sequence, predicts, seven,

```


Db	121	TGGGTCATAGAGCCCGGAAAGTGGAGAGCAGCTGCCCTGGAGGACCTTGACGTTTAC	180
Oy	724	Tgcccctcaagcgccatctcgcgaacgtaacgtaagcctggtgtacgcgcctcaacgycg	783
Db	181	TGCCCTACACGCGCATTCGGCAACGTCACGGGAGAGCTGGGTATCGGCCACTACGGGCGG	240
Oy	784	cccgaaagacctgcagagacctgcgcgcgaagggcggtgaatccagtgggcgcctgtctg	843
Db	241	CCCGAAGACCTGCAGAGACCTCGCGGCCAGGGGCGCTGGATCCAGAGGGCCGCTGCTGCTG	300
Oy	844	gtgcgcgtggggggtgatacagcttcgcgcgaaggtgacacatgtctcaagcttcggggt	903
Db	301	GTGCGCGGGGGGTATCAGCTTGCCCGCAGAAAGGTGACATGCTCAGGACTTCGGGGCT	360
Oy	904	caagagatgctcatataccagagcaagcggaactctccaggaaccacccaagccaaagt	963
Db	361	CAAGAGATGCTCATATACCCAGACCCAGCGACTTCTTCCAGGAGCCACCACCAAGCCAAAGC	420
Oy	964	ctgtccagccagcagcagtgtaatgtacatgtgcactgtggaactgtgagaccctacaca	1021
Db	421	CTGTCCAGCCAGCAGGAGGTATATGAGCAATGTGCACCTGGGAATGTGAGACCCCTTACAC	480
Oy	1024	ccgtgacttcctctccttcaatcaaaaccgaattccctcagtlgcatcatcaggcttccc	1083
Db	481	CCTGGCTTCCTCTCTTCAATCAACCAACCGAGTCCCTCAATGGATCATCAGGCTTCCC	540
Oy	1084	agcatccagcccaagcccat 1103	
Db	541	AGCATCCACGCGCGAGCCCT 560	

RESULT	4
LOCUS	GGTFFR 2500 bp mRNA VRT 30-SEP-1991
DEFINITION	G.gallus mRNA for transferrin receptor.
ACCESSION	X55348
NID	963806
VERSION	X55348.1 GI:63806
KEYWORDS	transferrin receptor.
SOURCE	chicken.
ORGANISM	Gallus gallus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
AUTHORS	Megnathae; Galliformes; Phasianidae; Phasianinae; Gallus.
TITLE	1 (bases 1 to 2500)
JOURNAL	Chan,L.N.L.
REFERENCE	Direct Submission
AUTHORS	Submitted (01-NOV=1990) Chan L.N.L., University of Texas Medical
TITLE	Branch, Department of Human Biological Chemistry and Genetics,
AUTHORS	Univ. Texas Medical Br., Galveston, TX 77550 U.S.A
TITLE	2 (bases 1 to 2500)
AUTHORS	Gerhardt,E.M. and Chan,L.N.L.
TITLE	Structure and Organization of the Chicken Transferrin Receptor and
AUTHORS	its cDNA Sequence
JOURNAL	Unpublished
FEATURES	
SOURCE	Location/Qualifiers
	1..2500
	/organism="Gallus gallus"
	/db_xref="taxon:9031"
	/cell_type="erythroblast"
	/clone_id="gt11.chicken erythroblast"
	/clone="CTRBV4"
	79..2415
	/standard_name="CTRCDNA1"
	/codon_start=1
	/product="chicken transferrin receptor"
	/protein_id="CAA39035.1"
	/db_xref="PID:963807"
	/db_xref="GI:63807"
	/db_xref="SPTRMBL:Q9097"
	/translation="MDHARALSNLFSVEPNSTYRSIAROTGDGNSVEMKLSADDE
	EGGIEPEHNAHVSMAOPNRNGRIKCLVIAAALLIGFLIGLISYRGHETLARCO
	DGSGGCITPTASTYLVDEGEGVEEIEGPPVTEFWPELKAMLSKLSAKNLYVDNLRFRV

GVSDFEAGAEADPNMAYTHIEEERNPLDKYMNDEHYTKLQVRCSTNOVISJINGKEE
 ILEPPEADIPYSESGSCKPVYVYNGKLKDFELIQKVASLNGITVYRAGKITLAE
 KVAARKDAGAGVLMYVDSLTMYGLTFLPILFEGHJHJTDGDPYPGFSEPHHTPPPE
 SSGPLHIAVOTISSASAARLTKSRMDQSGEGKGAHSCRTVKTKESQIMVLVDVN
 SMDKRLTINLFGLAQGFEEEDRVYVIGAQDSGKPGAKRGSTJATALLIELARISIV
 KNEBCKPRRSIIIFASWSAGDYGAVALIEWEGSAMIHAAFYIISIDAVPFLASHW
 ISAPLMLGLSJMKGVDKNPAVSESLYRNDGQVAVAPGLINAAFPPLAYSGI
 PLYSFGFYNDDEEYRFLDTGDTLENRKIDNDLALAAAEVAGAAALITJHDELF
 LDIGRYSEELLAQOEFLPYIKEVRELGLTLDMLFEALAGFORAVALARDINDSDE
 NRYRALRNDRMKMKVEYDFLSPYISPDEYFRIRIFCKGQFTLRISLVEHLQILKTNS
 SYVDLNLREQLALATWIKRGPRMPEYISCKLIMNSH"

BASE COUNT 650 a 545 c 707 g 598 t

ORIGIN

Query Match	10.6%	Score 305.4	DB 4	Length 2500
Best Local Similarity	52.2%	Pred. No. 1.2e-48		
Matches 833	Conservative 0	Mismatches 716	Indels 48	Gaps 5
QY 704	tggaagaccctgaagcgtctactctgcctccctcaacagcccatcgcaacgtctcaaggaagctg	763		
Db 737	TGGAACACTCTGATGACGATCTGTTGCATATACACGAGAGGGCTCTGTTCTTGCAACCTG	796		
OY 764	tgtaagcccaactaagggcgccgcaagaacctgcagagacttcgggccaagggcgctgac	823		
Db 797	TCTATGTGAACTACGGGCTGAAAAAGTTTTGGATCATACAAAGGCTCTGGCTTCAC	856		
OY 824	caatggcgccgctcgtctgctgctggcgctgggggtgatcaagcttcgcaccaaggtgaca	883		
Db 857	TGAATGAAACCATAGTACTTGTGTGACAGAGCTGGAAAAATATACACTTGCTGAGAAAGTTGCA	916		
OY 884	atgctcaggaacttcggggtcaagaagatgctcatataaccagagccagcggaactctcc	943		
Db 917	ATGCCAAGAGCGACGAGACACCTGAGAGTCCCATGTACGT-----	957		
OY 944	aggaccaccaccaagccaagcctgcacagccagccagcagtgatgagcaatgtgcactg	1003		
Db 957	-GGATTCACCTCATGTATGAGATTAACAGATACACTTATCCCTTTGGACATCCCACTTG	1015		
OY 1004	gaactggagaccctcacacacacttggctcccttccttaalcaaacccagttccctcag	1063		
Db 1016	GAACTGGAGACCTTATACACCCAGGCTTCCTTGTTCACACACACCAGTTCCACACAG	1075		
OY 1064	ttgcatacagggccttcgccagcatcccaagccagccatagtgagcaatttgcctccc	1123		
Db 1076	TTGAATCTTTCAGGACTACCCACATTTGCTGTTCAGACCATTTCTACAGTGCACACGCA	1135		
OY 1124	gctctgctgaagaaagctcaaaagccctcttgcgcccccaagaaatgycagggagccctc	1183		
Db 1136	GGCTTCTCAGCAAAATGATGTGAGACACATCTCTGAAGTTGGAAAGTGCATCATTT	1195		
OY 1184	gct-----ccccttatcacttggcccccgggccaagacttcggtgtgtcaaatc	1237		
Db 1196	CCTGTAAAGTACAAACAAAGAGGAGACCGAATPATGTAAACTAGATGTGAACAATT	1255		
OY 1238	acagagaccctcacccccctcaacaacacatcttcgctgcacatgaaagccgcctcaagccag	1297		
Db 1256	CCATGAAGACAGGAGAAATTCTGAACATCTCGGTGCTATCCAGAGATTGAACAACTG	1315		
OY 1298	atcaactagttgtacatcgvgggcccaagaggaatgcatvggggcccagaagcaagtaatccg	1357		
Db 1316	ATCGGTATGTTGTATATGTGAGCCCCAGAGAGACTCTCGGGGCCCGAGAGTGGCTTAACCTG	1375		
OY 1358	ctgtgggaagcgtacatactctctgtagctcgtgtgcggaacttctctccatggttgagcaac	1417		
Db 1376	GCACAGGAACGCTATATGTGTGGAACCTGGCCCTGTGATCTCAGACATATGTGAATAACG	1435		
OY 1417	--gacttcggccccgcagaagctctctcttcatacagctvggagcgtgtgtgactttgaa	1474		
Db 1436	AGGGCTAACAAACCAAGGCGAAGCATCATCTTTCCTACTGTGAGTGCAGGAGACTACGAG	1495		
OY 1475	gcgttggctcaaggaatgtgtgtagaaagctacccctcagcgttcttcacactcaaacgctag	1534		

Db	1496	CTGAGGCTGCTACGATGCGTGGAGGGGCTACTGCGCATGCTCATGCGCAAACTTTCA	1555
QY	1535	tgtagctgagccctgagacaagcagtgctgaggagtagacaagttcatgccaagaccagcc	1594
Db	1556	CTTACATCAGCTGGATGCTCCACTCTGGGAGACGACCATGATTAATTTCTCCAGCC	1615
QY	1595	ccctctgacaagctcatatgagagtgctcctgaaagcaggtgagatctcccaaccagtg	1654
Db	1616	CTTGTGCTATATGCTCTGGGAGATTAATGAAGGGGGTGAAATCCAGACAGCTCT	1675
QY	1655	gagcagactcctatgacaaggtggtgttcaaccaatcccaagctggagatgctgagtgacc	1714
Db	1676	CAGAGACCTCTATACAGACTTGGCCAGACTGGGTAAGAGCTGT-----	1726
QY	1715	gagccctaccacatgagacagcagtgccctatctcctcaagcccttgaggagctccg	1774
Db	1726	-CCTTTGGCTGATATGCTCTGGGAGATTAATGAAGGGGGTGAAATCCAGACAGCTCT	1783
QY	1775	tcgaggtcctctatgagagcagcagcagctccatctcctgacacaaagagagaca	1834
Db	1784	TGCTTTGTTCTTACAAATAAAGATAGAGAAATTCCTTCTGGACACTAAGGGTGACA	1843
QY	1835	cttatgagaacctgcatagaagtgctgcaagcgccctgcccgcgtgagccagccgtg	1894
Db	1844	CACGTGAGAACTGAGGAAATATGATAAT-----CTGGATGCTCTTCTGGCTCTCTG	1897
QY	1895	cccagctcgcagagcagcctcctcagcagcagcagcagcagcagcagcagcagcagc	1954
Db	1898	CAGAAAGTCTGACAGAGCTCTCAGGCTGAGCCATGATGATGATGATGATGATGATG	1957
QY	1955	tcggagcctcggagagagtgctgctcctcagcagcagcagcagcagcagcagcagc	2014
Db	1958	TGGGGAATACAGGAAATTAATGCTGATACAGGAGAGGATTTTCTTACATTAAGG	2017
QY	2015	acctcaagagccgcgagcagcagcagcagcagcagcagcagcagcagcagcagc	2074
Db	2018	AAGTGGGAGAGCTGGGTTGACTTGGACTGCTGTTTTCCTTCTGCTGCTGCTGCTG	2077
QY	2075	ggagcgcgcgagaaagctgagcagcagcagcagcagcagcagcagcagcagcagc	2134
Db	2078	GAGCTGTACTGCACTGAGAAAGACATGCAAACTGAGGAGGAGGAGGAGGAGGAGG	2137
QY	2135	caagcagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2194
Db	2138	GCAAGGCCCTGAATGACAGATGATGAGGATGATGATGATGATGATGATGATGATG	2197
QY	2195	cgccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2254
Db	2198	CACCAAAAGATGTCCTTTTCGCCACATCTTTTGGGCAAAAGGCCCCACACCTGCGGA	2257
QY	2255	ccctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2291
Db	2258	GTCGTGTGAGAGATCTGACAGCTGTTGAAGACACAGAG 2294	
RESULT	5		
HSTR			
LOCUS	5010 bp	mRNA	PRI
DEFINITION	Human mRNA for transferrin receptor.		11-APR-1995
ACCESSION	X01060		
NID	337432		
VERSION	X01060.1 GI:37432		
KEYWORDS	transferrin receptor.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;		
AUTHORS	Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 5010)		
JOURNAL	Schneider, C., Owen, M.J., Banville, D. and Williams, J.G.		
MEDLINE	Primary structure of human transferrin receptor deduced from the		
	mRNA sequence		
	Nature 311 (1987), 675-678 (1984)		
	85012743		

COMMENT	Data kindly reviewed (19-FEB-1986) by C. Schneider.	
FEATURES	Location/Qualifiers	
source	1..5010	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
CDS	264..2546	
	/note="put. transferrin receptor (aa 1-760)"	
	/codon_start=1	
	/protein_id="CAA25527.1"	
	/db_xref="PID:g37433"	
	/db_xref="GI:37433"	
	/db_xref="SWISS-PROT:P02786"	
	/translation="MMDQASAFESNLGGEPLSYTFRSLARQVDGNSHVEKMLAVDE EENADNNTKANVTQPKRSGSISICGTIAIVVFLLIGFMIGLYCQGVPEKCEERLA GTESPVREPEGEFPAARLYMDLKRKSEKIDSDSTSTIKLHNSVYPREASQ KDEMLALYENOFREFRLSKVMDQHVVKIOYDSQNSVVIYDKNGRLVYLVEENG YVAASKAATVTKPLVANGTKKDFEDLTTPNGSIIVIRAGITFPAEKVANSINA IGVLIYMDOTKPIYNAELSFCHALIGTGDPTTPGPFSPNHTQFPSSRSGCPNIPV OTISRAAEKLFQNEBDCPSDKTIDSTQRMVTSSEKNVLTIVSNLKEIKLINIFGV IKGEVDPDHYVAVAGQADMBGPAASGVGTALLLKLAQFSDMVLDKGFQPSRIIF ASWSADGFSVGAATEWLEGLSLHLKATYINLDRAVIGTSFKVASPLATYLLIEK TKQNTVHPVTGQFLXODSNMWSKVEKLTIDNAAPFLAYSGIAPVFCEDPDYPL GTMQDYKLEIRIPELNVVRAAFAVAGQFVYKTHVDENLDYRYSOLSEFVRD LNQYRADIREKGLSLQMLYSARGDFPRAISRLITTDGNAEKTRFPMKLNDRMAYVE YHPLSPVSKESPFRHVFWSGSHLPLALELKLKRONNGAFNETLFRNOLATMT TIQGANALSGVDWIDINEF"	
	456..527	
	/note="put. transmembrane segment (aa 63-88)"	
	1014..1022	
	/note="pot. extracellular glycosylation site (aa 251)"	
	1212..1220	
	/note="pot. extracellular glycosylation site (aa 317)"	
	2442..2450	
	/note="pot. extracellular glycosylation site (aa 727)"	
BASE COUNT	1399 a	977 c 1155 g 1479 t
ORIGIN		
Query Match	9.8%; Score 280.8; DB 9; Length 5010;	
Best Local Similarity	54.1%; Pred. No. 9.4e-37;	
Matches 781; Conservative	0; Mismatches 607; Indels 56; Gaps 8;	
QY	839	tgctgtgctgagcagtgagtgatcagcttcgcccagaagtgacaatgctcagagctcg 898
Db	1027	TGATTGTCAAGACGACGAAATACCTTTCCAGAAAAGTTCGAAAGCTGAAGCTTAA 1086
QY	899	gggtcagagagtgctatatccagagcagcagcagcagcttcccagagaccaccaagc 958
Db	1087	ATGCAATTTGGTGTGTGATATACATGAGCAGCACTAATTT-----TC 1128
QY	959	caagcctgctcagcagcagcagtgatgagacatgagcagcagcagcagcagcagc 1018
Db	1129	CAATTGTTAACGACAGACTTCAATCTTTTGGACATGCTACTTGGGACAGTGACCTT 1188
QY	1019	acacacctgctccctcctcctcaatcaaaaacacagttccctcagttgcatcagcc 1078
Db	1189	ACACACCTGATGATCCCTTCCTCAATCATCACTAGTTCCACCATCTCGGTCATCAGAT 1248
QY	1079	ttccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1138
Db	1249	TGCTTAATTAATCCTGTCACACAAATTCACAGACTGCTGAGAAAAGCTGTTGGGATA 1308
QY	1139	tcaaaag-ccctgtgagccccaagaatgagagagcagcagcagcagcagcagcagc 1197
Db	1309	TGGAAGAGAGACTGTCCTCTGAGATGAGAAACAGACTTATATGATGATGATGATGATG 1368
QY	1198	ctgggcccgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1257
Db	1369	AAAGCAAGAATG-----TGAAGCTCAGCTGTGAGCAATGTGCTGAAGAAAGATTAAT 1421
QY	1258	aaacaacatctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1317
Db	1422	CTTAAATCTTTGGAGTATTAAAGGCTTTGTGTAACCAAGCATCATATGTTGAGTTGGG 1481

```

OY 1318 gcccaaggatgcatgagggcccaaggagcagcgaatcccgctgtgagagcgtatacc 1377
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1482 gcccaagagatgcattgagggcccttgagctgcaaaatccgggtatgacacagctctccta 1541
OY 1378 ctgagagctgtgcgagaccccttccctccatg--tgaagaacgagctccggcccccaga 1434
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1542 tggaaacttgcccaagatggtctcagatgctttaaagatgggtttcagccacacaca 1601
OY 1435 agtctcccttcaatcagctcgagagctgtgtgactttggaagcgtggtgccagagctg 1494
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1602 accattatcttggccaggttgagctgtgagactttggatcggttgcctcactgatatg 1661
OY 1495 ctgaagagctaccctcagctgtgcacccccaagcgtagtgaagctgagcccgagcaac 1554
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1662 ctgagaggtaccttcttgcctcctgcctttaaaggctttacttatattatctggatpaa 1721
OY 1555 gcaagtgtgagggatgacagattcactgcgaagaccagccctctcgaagctcatt 1614
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1722 ggggttcttggtacacgcaacttcaaggttttgcagcccaactggttatagcgttat 1781
OY 1615 gagatgtctctgaagcaggttgatctcccaaccagctgggcagactcttatgacag 1674
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1782 gggaaacaatgcamaaatgtgaagcattcc--ggttactggcgcaatttctatcagaga- 1838
OY 1675 gtggtgtcccaatcccaatccgagtgagtgagtgatcggccctcaccatgagcagc 1734
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1838 -----cagcaactggccacgaaagttgagaaactcctttagacatt 1880
OY 1735 agtgcctatctcctcagcgccttgtgagagtcctcgtcgagctccttctatgag 1794
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1881 gctgttcttcccttcttctcctatattctggaatccacagatttcttctgttttgcag 1940
OY 1795 gacgcacagcctaccatctctctgcacacaaagagagacattatagaactgtcaag 1854
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1941 ---gacacagattatcttatttgggtaccacacantgacacactatgaagacattgag 1997
OY 1855 gtgcgcaagcgccgctgcgcgcgtgcgcgcagtgccagctcgaagcgcaagcgc 1914
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1998 aggatttccg---agttgaacaaagtgccacagcagcctgacagagctcctgtcatttc 2054
OY 1915 ctcatcgcgcctcagcagcagatcgctgtcgcctcgcactcgcgcgcgtacagggagcgc 1974
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2055 gtgatttaacttaaccatgatgttgattgaactcgactgagatgagaggtacaacaccaa 2114
OY 1975 gtccctcagcacaatcgaggaaactcaacgagttctctgggagacctcaagcccgagctg 2034
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2115 ctgcttttcatTTGTGAGGGATCTGAACCAATACAGACAGACATAAAGAAATGGGCTG 2174
OY 2035 accctgcagtggtgtgactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2094
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2175 agtttacaatggcgtgatttctgctgctgagacatttttccgttcttccacacttaaa 2234
OY 2095 cagagagatctacagctcggagagagagagacgagcgaactgacacgcatgacacgctgcgc 2154
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2235 acagatttgcggaatgctgagaaacagacagatttgcatagaagaactcaatgatcgt 2294
OY 2155 ataaagcggtgtgagttcacttcttccagtaagtggtgcgaacgcgaactcccgcttc 2214
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2295 gtcatgagagtgagatgacttcttcttcccttcccttcccttcccttcccttcccttcc 2354
OY 2215 cgcacatctcatgagcgctgagagacacacagcgtgagcgctcgtgagacacacgcgcgcg 2274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2355 ccacatgcttcttggggctccggctctcacaacgcttccacgcttcttcttcttcttctt 2414
OY 2275 ctgc 2278
    |||||
DB 2415 ctgc 2418
    |||||

```

```

RESULT 6
HUMTERR HUMTERR 2826 bp mRNA PRI 14-JAN-1995
LOCUS

```

```

DEFINITION Human transferrin receptor mRNA, complete cds.
ACCESSION M11507
VERSION 9339515
KEYWORDS M11507.1 GI:339515
SOURCE Human, cDNA to mRNA, clone pCDTR-1.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS 1 (bases 1 to 2826)
McLellan,A., Kuhn,L.C. and Ruddie,F.H.
TITLE The human transferrin receptor gene: genomic organization, and the
complete primary structure of the receptor deduced from a cDNA
sequence
JOURNAL Cell 39 (2 Pt 1), 267-274 (1984)
MEDLINE 85048936
COMMENT Draft entry and clean copy sequence for [1] kindly provided by
A.McLellan, 17-Apr-1985. The area coding for the transmembrane
domain of the transferrin receptor protein is found at positions
279-362.
FEATURES
source
location/Qualifiers
1..2826
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="3q26.2-qter"
<1..>2826
/note="tfr mRNA"
96..2378
/gene="TFRC"
96..2378
/note="TFRC"
/gene="TFRC"
/note="transferrin receptor"
/codon_start=1
/db_xref="GDB:600-120-433"
/protein_id="AA61153.1"
/db_xref="PID:9339516"
/db_xref="GI:339516"
/translation="MMDQARSFNLFGEPILSYTPRESIAQVQDQNSHYEMKLADE
EENADNNTKAVNTTPKRCSSGTCGTAIVYFLIGMIGTLGCKVEPTDEBRLA
GTESPVEPEDEPFAARKLYWDLKKRLSKLDSTPTSTIKLNNESVYPRRSGO
KDEMLALYVENQREFKLSKWRDQHEVKIQVDSQNSVLIIVDKNRLVYLVENPG
YVAASKATVYGVKLVANFGKDFEDLYTPVNGSIVYVRAGITFAEKVNAESLNA
IGVLIYMDOTKEPLVNAELSPFGAHIGTDPYTPGPESEHQTQEPSSRSGLENIY
OTISRAAEKLFQMEGDDCPDMDKTDSTCRVSESNKLYTVSNVLYKEIKLIMFGY
IKGVEPDHYVYVAGQRDAMPGPAKSGVGTALLIKLAQMSDVLKLDGFPSSITF
ASWAGGFGSGVATGEMLEGLISLHLAFYIINDKAVLGTSNKVSASPLLYLIEK
TMQNVKHPVTGOFLYQDSNMAKVEKLTLDNAAPFLAYSGIIPAVSFCFBDYPLY
GTTMDTYKELIERLPELNVKVARAAEVAAGQFVIRLTDVELNDLYERNQSLSFVD
LNOYRADIKEMGLSLQWLYSARGDFEATSRITDFGNAEKTDREVMKILMDRYMRE
YHFLSPVSPRESPEFRHVFGSSHTLPALLEMLKLNKNGAFNELLFRNQLALIV
TIGQAAALSDVDIDNEF"
BASE COUNT 793 a 568 c 676 g 789 t
ORIGIN 449 bp upstream of PstI site.
Query Match 9.7%; Score 280.2; DB 10; Length 2826;
Best Local Similarity 55.0%; Pred. NO. 1.3e-36;
Matches 717; Conservative 0; Mismatches 348; Indels 38; Gaps 7;
OY 980 cagttatggaacttgacactggaacttgagaccctacacacactgctccctct 1039
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 982 catttcttggacatctctatcctggggacagcagacccttaccacacttgattcccttct 1041
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1040 tcaatcaaacccagttccctccagttgacatcagccttccagcatcccaagcagc 1099
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1042 tcaatcacacacagtttccaccatctcggtcatcagattcgctaatattacctgacaga 1101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1100 ccatcagtcagacattgctcccgccgctgcgaggaactcaagc-ccctgtgcccc 1158
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1102 caatctccagacctgctcagaaaagctgttgggaanaatggaagagacgtgtccctctg 1161
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1159 caagaatgacagggagcctcctcagcctcccttaccacctcgagcccgcccaagactg 1218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 1162 ACTGGAAAAACAGACTCTACATGTAGATGGTAACCTCAGNAAGCAAGATG-----TG 1214
Oy 1219 cggcagtagtgaacaacacacagagcccccacacacacatcttgcgtgcatc 1278
Db 1215 AAGCTCACTGTGAGCAATGTGCTGAAGAGATAAAAATTTCTTAACATCTTTGGACTTAT 1274
Oy 1279 gaagccgcctcagagcagatcactacgtgtgcatcgagccagagagatcaggggc 1338
Db 1275 AAGGCTTTGTAGAACAGATCATTATGTGTAGTTGGGCCACAGAGATCATGGGCC 1334
Oy 1339 ccaggaagcagcctaactccgcgtgtgggagcagcctaactctcgtgagctgtgcgagcc 1398
Db 1335 CCGAGAGCTGCAAAATCCGGTGTAGGACAGCTCTCTATTGAAACTTGCCAGATGTC 1394
Oy 1399 tccctcaggg---tgagcaagcgttccggcccgagagatcctcctcctcagctgc 1455
Db 1395 TCAGATATGGTCTTAAAGATGGGTTTCAGCCACAGAGAGATTAATCTTTGCCAGTTGG 1454
Oy 1456 gacggtgtagcttggaaagcgtggggtccacagagtgagtgaaagctacccagcgtg 1515
Db 1455 AGTGTGGAGACTTTGAGATCGGTGTGTCACACTGAATGGCTAGAGGATACCTTTCTGCC 1514
Oy 1516 ctgcaccccaagccagctgtgtagcgtgagcctcgaacagcagctgtggtggagtagcaag 1575
Db 1515 CTGCAATTAAAGCTTTCACTATTATTAATCGATTAAGCGGTTCTTGATCCAGCAAC 1574
Oy 1576 ttctctgcaagaccagcccccctctcgaagatcctcatgagagctgtccctgaagcagtg 1635
Db 1575 TTCAAGGTTTCTGCGACGCCACTGTTGATACGCTTATGAAACAAATGCAAAATGTG 1634
Oy 1636 gattctcccaacacagctgggagcagctcctctagacaggtgtgttccacatcccaag 1695
Db 1635 AAGCATCCGGT---TACGTGGCAATTTCTATATCAGA-----CACG 1673
Oy 1686 tggagctgtgagtgatcccgccctcaccatggaacagcagtgctatcctcctcagcgc 1755
Db 1674 AACTGGCGCAGCAAGTTGAGAACTCACTTACAGCAATGACGCTTCCCTTCCCTTGA 1733
Oy 1756 ttctgtggagagtcctccgcgtgagtgctcctcttctgaggaagcagccagcctacatc 1815
Db 1734 TATTTGGAATCCAGCAGTTTCTTCTGTTTTCGGA--GACACAGATTATCTTAT 1790
Oy 1816 ctgcacacaaagaggaacactatgagaacctgacaaagtgtctgcaagccgcctgccc 1875
Db 1791 TTGGGTACCACTGATGACACCTATATAGGAACTGATGAGAGATTCTG--AGTTGAAC 1847
Oy 1876 gccgtggcccaagccgtgtgcccagctcgaagcagcagctcctcctcgcgtcagcagcag 1935
Db 1848 AAAGTGGCAGCAGAGCTGACAGAGGTGGTCAAGTTGATTAACATAACCATATAT 1907
Oy 1936 cgcgtcgtccctcgaactcggcgcgtacgagggagcgtctcctcgaagcagcagcagc 1995
Db 1908 GTTGAATGAACTGACTGATAGAGGTACAACAGCAACTCTTTCATTTGTGAGGGAT 1967
Oy 1996 ctcaacagagttctctgggagacctcaagcccgagcgtcagcctcagtggtgtactcg 2055
Db 1968 CTGAACCAATATAGAGCAGACATTAAGGAATGGGCTGAGTTTACAGTGGCTGATTTT 2027
Oy 2056 gcgcgggggagactacatccggcgcgcggaagcgtgcgcgcaagagatcctcagcgcgcgag 2115
Db 2028 GCTGTGAGACTTCTCCGCTGCTACTTCCAGACTAACACAGATTTCGGGAATGCTGAG 2087
Oy 2116 gagagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2175
Db 2088 AAAAGACAGATTTGTATGAAGAACTCAATGATCGTGTATGAGATGAGATGATC 2147
Oy 2176 ttcccttcagtagctgtgcgcagcagcagcagcagcagcagcagcagcagcagcagcagc 2235
Db 2148 TTCTCTCTCTCCCTACGATATCTCAAAAGAGTCTCTTTCCGACATGTCCTTCTGGGGCTCC 2207
Oy 2236 ggaagacacagcgtgagcgcctctgtgagccacactgcgcgtgc 2278
Db 2208 GGCTCTCAACAGCTGCGACCTTACTGAGAGAACTTGAACCTGC 2250

RESULT 7
CRUTRANREC CRUTRANREC 2460 bp mRNA ROD 09-DEC-1993
LOCUS Cricetulus griseus transferrin receptor mRNA, complete cds.
DEFINITION L19142
ACCESSION L19142
VERSION g304528
KEYWORDS L19142.1 GI:304528
SOURCE transferrin; transferrin receptor.
ORGANISM Cricetulus griseus cDNA to mRNA.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
REFERENCE 1 (bases 1 to 2460)
AUTHORS Collawn,D.F., Lai,A., Domingo,D., Fitch,M., Halton,S. and
Trowbridge,I.S.
YRRF is the conserved internalization signal of the transferrin
receptor, and a second YRRF signal at position 31-34 enhances
endocytosis
J. Biol. Chem. 268 (29), 21686-21692 (1993)
94012749
FEATURES
MEDLINE
SOURCE Location/Qualifiers
1..2460
/organism="Cricetulus griseus"
/db_xref="taxon:10029"
/cell_line="CHO-K1"
/tissue_type="Ovary"
/tissue_lib="Stratagene catalog #937910"
1..2274
/codon_start=1
/function="endocytosis"
/product="transferrin receptor"
/protein_id="AA03576.1"
/db_xref="PID:g304529"
/db_xref="GI:304529"
/translation="MMDQARSALSNLFGEPSTYRPSLIAROVNDGSHVEMKLAVDE
EENTDNMKASVKKRRLNGLRCGTIAVYIFLIGRMIGLIGCKRTBOCDYRLAE
TETGNSLTIQENIPQSSRLYVADLKLLEKDLAIEFTDIKOLDSOTSRAGSKDE
NLAYIENQFDEFLSVWREHYVKIQVKSAAQNAVTIINVGSDSLVENPGYVA
YSKATVYSGKLIIHNFGLKDFEDLKFPVNSLIVRAGKITPEKVAANQSFVAIG
LYVDOTKPPVVEALSLFGHAIIGTDDPYTGPFPFNFHQPSPSSGSLPSIVOTI
SRKAEKLQNFMEPCPSNMWDTCTLESQGIINVLNVNVIKRRILNFEVIG
FEEDRITVGAQDANGPGNAKSVCTGLLKLQAFSDKVSAGRPSSIIIFASV
SAGPQGVAGATEMEGLSLSLHLKAFYINLKVLTGRNKSVASPLLYLIEKTMQ
DVRHPIDGKPLVRDSNMIISKVEDLSLNAFPLFAYGIPAVSWPCENEDYPLDTN
LDYREKLQKVPOLNKKWRAAEVAGQFIKFLHDIELNDYDMYNNKILISVELNO
FRADIKAMGSLQWLYSARGDFPATSRLITDFPNARKNTNRFVYRITNRMKVEYH
LSRIVPSRESRFRIIFWSSGSHITALYENLKLKKNSSAFNETLFRQALATWTTIO
GVANALSGDILWDIDNEF"
2275..>2460
BASE COUNT 744 a 487 c 539 g 690 t
ORIGIN
Query Match 8.9%; Score 256; DB 12; Length 2460;
Best Local Similarity 54.2%; Pred. No. 1,16-32;
Matches 713; Conservative 0; Mismatches 565; Indels 38; Gaps 8;
Oy 980 cagtgatggacatgtgacactgggaactggagaccctaacacactgtcttccct 1039
Db 878 CACTCTTTGGACATGCTCATCTAGGACCTGGGATGCCATACACCTGGGTTTCCATCTT 937
Oy 1040 tcaatcaaacacagttccctcagttgatacatcagagccttccacagcattcccaagccagc 1099
Db 938 TCAATCATATCAGCTTCTCCATCTCCATCATCGATGATGATGATGATGATGATGATGAT 997
Oy 1100 ccatacgtcagacatgtccctccgcgtcgaggaagct-caaagccctgtgagccccc 1158
Db 998 CGATCTCAAGAAAGGCTCCAGAAAGACTATTTCAAAACATGGAACAAACATGCTCTCCTTA 1057
Oy 1159 caagaatggcagggagagcctcctagagctccctatacctcgggcccggccagcagatg 1218
Db 1159 caagaatggcagggagagcctcctagagctccctatacctcgggcccggccagcagatg 1218

Db 1058 GTTGACACAGATTTCTTGTAGCTAGTAACTTCA-----CAGGGTATAATGTG 1110
 QY 1219 cggtagtgcgaacaatacaagaactcccccatacaacatcttgcgtcgcac 1278
 Db 1111 AACCTCTCTGTGAACAAGTCTGAAGAAACAAGAACTTAACATCTTTGGAGTTAT 1170
 QY 1279 gaagcgctcagagcagatcactacgtgtgtcagggcgccagagagatcgtgggc 1338
 Db 1171 AAAGATTGTGAGAACCAAGACCGGTACTTACTAGTGGCCAGAGAGATGCTGGGGT 1230
 QY 1339 ccagagcagactaaatccgcgtgtgtggagcggctactcctgtgagctgtgcgacctt 1398
 Db 1331 CCTGCTCTGTGAATAATCAGTGTGGAAACAGGCTTCTGTGAACCTTGCCCAAGCATTC 1290
 QY 1399 tccccatgfr---gagcaaggtctccggcccgccagaaatctcccttataagcttg 1455
 Db 1291 TCAGATATGTTTCAAGAGGTGGGTTTAAACCCAGAGAACATTAATCTTTGGCAGCTGG 1350
 QY 1456 gacggtgtgacttggaaagcgtgtgcctcagagagtgtagaaggctacatcagcgtg 1515
 Db 1351 AGTCAGAGAGACTTCGAGAGCTGTGTGTCACAGAGTGGCTAGAGGATTAATCTTCACT 1410
 QY 1516 ctgacactcaaaagcagtagtgcgtgagccttgagcaacagcagctgtggggagtagaag 1575
 Db 1411 CTGCATTGAAAGCTTCACTTATTAATCTGGATAAAGTTGTCCTTGTGTAAGAAC 1470
 QY 1576 ttcatgccaaagacagccccctctcgaacgtctcatitgagagtgctcctgaagcagtg 1635
 Db 1471 TTCAGGTTTCTGCGACCCCTCTATTAATACCTTAATGAGAGACAAATCCAGAGCTA 1530
 QY 1636 gattctcccaacacagtgagcagactctatgaacagsgtgcgttcaacatccagc 1695
 Db 1531 AGAATCCAAATGA---TGGAACCTCTCTATGAGA-----CAGAAATGGATC 1578
 QY 1696 tggagtcgtgagtgatccggcccttaaccaatgagcagtgctatcttccagcggcc 1755
 Db 1579 AGCAAACTTGAGG-----ATCTTCCCTTGACAAATGCTGCTTCCCTTCTTGCA 1629
 QY 1756 ttgtggagagccctgcgtgcgtgagttctccttatagagagcagcagccttaccatc 1815
 Db 1630 TATCTGGAATCCAGAGAGTTCTTCTGTGTGTAAGATG---AGGATTAATCTTAT 1686
 QY 1815 ctgacacaaagagagacactatgagacactgacataagagtgctgcgaagcgccctccc 1875
 Db 1687 TTGACACTATTTGATTAATCTAGAGAACTAATTCAGAAAGTTC---CTCAGCTGAC 1743
 QY 1876 ggcgtggcccaagcgctgtggccagctgcagggcagctcctcatccgcgtcagcagat 1935
 Db 1744 AAAATGTTGTCGACGACGAGAAAGTGTGCTGATTAATAAATTAACTTACCAGTAC 1803
 QY 1936 cgctgtgtccctgcagcttcgcgctlaagggagagtgctccctcaagcaaatcgggaac 1995
 Db 1804 ATTGAATGAACTGTGACTATGATATGATMAACAAGATAGTGTATTTGTGAAGGAG 1863
 QY 1996 ctcaacagatctcttggagacactcaagccccggcgagctgcagccccgagtggtactgc 2055
 Db 1864 CTGACCAATTCACAGCAGATATCAAGCAATGGGCTGTGAGTCTACAGTGTCTATTTCT 1923
 QY 2056 ggcggtgggagactacatccggcgggcgaaagactgcggcagagatcttaagctggag 2115
 Db 1924 GCTGTGTGAGACTTTTCCGCTACATCTAGACTGACGACTGATTCATTAATGCTGAG 1983
 QY 2116 gagggagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2175
 Db 1984 AAAACAACAGATTTGCTGTGAGGGAATCAATAACCGTATTAAGAAAGTGAATATAC 2043
 QY 2176 ttcccttcacagtagtgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2235
 Db 2044 TTCTGTGACCCCTACGATATCTCCAGAGAGTCTCTTCCGACACATCTTCTGGGGTTC 2103
 QY 2236 ggaagacacagcgtggcgccctgtgtgacaacactgcggtgtgtgtgtgtgtgtgtgtgt 2291
 Db 2104 GGCTCTACACTCTACAGGCTTTAGTGTGAGAACTTGAAGCTTCTGCTCAGAAAAACAG 2159

RESULT 8
 RATTFR 3413 bp mRNA ROD 27-APR-1993
 LOCUS Rat transferrin receptor mRNA, 3' end.
 DEFINITION
 ACCSSION
 NID 5207463
 VERSION M58040.1 GI:207463
 KEYWORDS
 SOURCE Rat adult 120 day old testis, cDNA to mRNA.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 3413)
 Robertis,K.P. and Griswold,M.D.
 Characterization of rat transferrin receptor cDNA: The regulation
 of transferrin receptor mRNA in testes in Sertoli cells in culture
 Mol. Cell. Endocrinol. 14, 531-542 (1990)
 FEATURES
 source
 1..3413
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /dev_stage="120 day old adult"
 /tissue_type="testis"
 1..1870
 /gene="transferrin receptor"
 <1..1870
 /gene="transferrin receptor"
 /product="transferrin receptor"
 /codon_start=2
 /protein_id="AAA42273.1"
 /db_xref="PDB:207464"
 /db_xref="GI:207464"
 /translation="TEFTDILKQLSNTYTPREAGSOKDENLAVYIENLHDFKESKY
 WRDEHYVKIOVKNVSQNLVTINSNSIDPVAPRGVAFSGAEVTKLVANSGTK
 KPEELNLYVNSLSIVRAGITLPAEKVANAQSPNAGVILYMDRNTFPVADLOF
 GHAGTGDPTPGPSPFHPNPOPSOSGLSIPVORTSAPAEKLEFMNMGNDPS
 WNIDSSCKLELSQNONVKLTNNVNLKETIILIFVYIKTEEPDRIVGADRDAMG
 GVAKSSVGTGLLKLAAOVFSDMIISDGFPSRSIIIPASWTADYGAVGEVLEGLS
 S.LHLAFYINIDKLVLTGNSFNKVSAPLTYLMGKIMQDVHPIDGKLYRNSNWS
 KIEELSDNAAPFLAYSGIPAVSFCECEDPYLGTLDYELIIOKVPOLNOMVR
 TAAEVAQGFILTFHDIETLDYEWNSKLSFMDLNOFRADIDMGSLQMLYSAR
 GQVFRATSLTDFDHFNAEKTNRFVAREINRKLIMKYDHLTGLSYVSPRESFPHITMS
 GSHTLSALVENLRLRKNITAFNETLFRNQLALATWTIGVANAUSGDIMWNDEP"
 2846..3007
 /note="putative iron regulatory region"
 misc_feature
 3184..3301
 /note="putative iron regulatory region"
 BASE COUNT 971 a 668 c 741 g 1033 t
 ORIGIN
 Query Match 8.6%; Score 247; DB 12; Length 3413;
 Best local Similarity 52.6%; Pred. No. 2.9e-31;
 Matches 767; Conservative 0; Mismatches 625; Indels 65; Gaps 8;

QY	1220	gactagatggtcaacacatcaaggactccaccccccaataaacaactctggtcgatcg	1279
Db	708	AGCTCACTGTGAACAAATGTAAGTAAGAAACAAAGAAATCTTAACATCTTGTGGCTTTATTA	767
QY	1280	aagtcgcctcaagagccagatcatcactagctgtgtcatcggggcccaagagatcatcggggcc	1339
Db	768	AAGGCTATGAGAAACCCAGACCCGTACATTTGTAGTGAAGACCCAGAGAACGCTTGGGGCC	827
QY	1340	caagagcagctaaatccgcctgttgggagcgctatactccctgagcttgggtgagacctttt	1399
Db	828	CTGGTGTGGGAAGTCCAGTGTGGAAACAGGCTCTGTGTGAACCTTCCCAATATTCT	887
QY	1400	ccctcatggtc---gagcaaggtctccggcccccgcgaagaatctcctcttcaatcgctgg	1456
Db	888	CAGATATGATTTCAAAAGATGATTAAGCATCCAGACAGAGATTAATCTTTGCCAGCTGGA	947
QY	1457	acgattgtgactcttggaaaggtgtggtctccacgagtggtgctaaaggtcactcaagctgc	1516
Db	948	CTCAGAGAGACTATGAGCTGTTGGTCCACACTGATGGCTGAGGGGTACTTTCACTT	1007
QY	1517	tgcacctcaagccgtagtgtacgttgaagcttggacaacgcaagtgtctggggatgacaagt	1576
Db	1008	TGCATCTAAAGCCTTTCACCTACATTAATCTGGATTAAGTCTCTGGGTAAGTCACT	1067
QY	1577	tctatgcacaagacagcccccctcttgcacaagtcctatctatgagagtgctcttgaagaagtg	1636
Db	1068	TCAAGTCTTCTGCAAGCCCCCTATTAATACATTATGGGAGAGTAATGAGAGACTTAA	1127
QY	1637	attctcccaacacacagtgggcgagacccctctatgaaacagtggtgttaaccaatccagct	1696
Db	1128	AGCATCCGATTGA---TGGAAATATCTATATCGAAA-----CAGTA	1166
QY	1697	ggagatgctgagtgatccggcccccctacccaatggacaagcatgctattccttcaacggct	1756
Db	1167	ATTGATTATCAAAATATGAGAACTTCTCTGGACAAATGCTCATTCCTTTCTTGAT	1226
QY	1757	tctgtggagatccctctgcgtctgaagttctccttatatgagagcaccaggtcctaccatcc	1816
Db	1227	ATTGAGAAATCCAGCAGTTCTTCTTGTGTTTGTGAGATG---AGGACATATCTTAT	1283
QY	1817	tgcacacaagggaggaacatctatgaaacctcataagtggtgtgtaagggcgcccttgcgcg	1876
Db	1284	TGGCACTTAACATGATACCTATGAAATTTAATTCAGAAATG---CTAGCTCAAC	1340
QY	1877	ccgtggccacagcgcttggcccagctctgcagggcagctctctcaatccgctcagcaagtc	1936
Db	1341	AAATGTTTCGTACAGCAGCAGAGGTGGCGGTCACTTATTTAATTCACCATGACA	1400
QY	1937	gctgtgcctccctcagcttgcgcgtacaggggagcgtgcctcagggacatcgggaacc	1996
Db	1401	TTGAGTTGACCTGGAGACTATGATGATTAACACGACAACTACTGTCATTATGAAAGATC	1460
QY	1997	tcaacaggttctctcggtgacctaaagcccggggtctgaccccttgaaagtgggtgtactcgg	2056
Db	1461	TGAACCACTTCAAGCAGATATTAAGATATGGGTCTAAGTCTTAACAAAGCGTGATTCGTG	1520
QY	2057	cgcgggggagcatcatccgggcggcgcggaagaagctgcgagagagatctacaagctcgaggg	2116
Db	1521	CTCTGGAGACTACTTCCGTGCTACTTCTTAAGACTAACACTATTTTCATATGCTGAGA	1580
QY	2117	agagagacgagcgacttgcacacgcatgtacaacgttgcgtacaaatgtcgggttgaatttact	2176
Db	1581	AAACAAACAGATTTCGTATGATGAGGAATCAATGATGATATTTGAAATGGAGATCACT	1640
QY	2177	tcccttccagtgatcgttgcgcagccgactccccgttccgcacaatcttcaatggtccgtg	2236
Db	1641	TCTCTACCCCTATGATATCTCAAGAGAGTCTCTTCCGACACATCTTCCGGGCTCTG	1700
QY	2237	gagacacacagcttggcgcccttgcagacacacttgcgcttgcgtcccaacagctcgcg	2296
Db	1701	GCTCTCACACCTCTCAGCTTTAGTGGAGAACTGAGACTTGTCTCAGAAAAATAT-----	1756
QY	2297	ggaccccccggggccactctctcaacttgcttccagagagccggttccggcgctcaagctag	2356

	Dd	1756	-----CAGTCGCTTTCAATGAAGAAGCGCTTCCACAAACCAGCTTG	1793
Oy	2357	cacctctaccctgacgtcgaaggggcagcccaactcgcttagcgaggatgctcgaaca	2416	
Dd	1794	CCCTGGCTACTGTGCATATTACAGGAGATGCCAATAATGCCCTCTGTGTGACATTGGAA	1853	
Oy	2417	ttagtaaacactctga	2433	
Dd	1854	TTCACAATGAGCTTTAA	1870	
RESULT	9			
LOCUS	MNTRRMRNA	2292 bp	mRNA	
DEFINITION	M.musculus	RNA for transferrin receptor.	ROD	
ACCESSION	X57349		29-OCT-1992	
NID	G54914			
VERSION	X57349.1	GI:54914		
KEYWORDS	transferrin.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;			
	Rodentia; Sciurognathi; Muridae; Musinae; Mus.			
REFERENCE	1 (bases 1 to 2292)			
AUTHORS	Trowbridge,I.S.			
TITLE	Direct Submission			
JOURNAL	Submitted (23-JAN-1991) I.S. Trowbridge, Salk Institute, P O Box 85800, San Diego, CA 92186 5800, U S A			
REFERENCE	2 (bases 1 to 2292)			
AUTHORS	Trowbridge,I.S., Domingo,D.L., Thomas,M.L. and Chain,A.			
JOURNAL	Unpublished			
FEATURES				
source				
CDS				
	Location/Qualifiers			
	1..2292			
	/organism="Mus musculus"			
	/strain="C57BlxB/2"			
	/db_xref="taxon:10090"			
	/tissue_type="hematopoietic"			
	/cell_type="70z/3"			
	1..2292			
	/codon_start=1			
	/product="transferrin receptor"			
	/protein_id="CAA40624.1"			
	/db_xref="PID:g54915"			
	/db_xref="GI:54915"			
	/db_xref="SPRMBL:O62351"			
	/translation="MMDQARSAPSNLFGEPELSYTRFSLARYVDGSDNSHYEMKLADE EENADNNMKASVKKPKREFNGRLCFALIALIVFLIGMSGYGICRKVEOKEECVKLA ETELTDKSETMETEDVPITSSRLYMADLIKLLEKNSIEFADTIKQLSQNTYTPREG SQDESILAYILENOHFHEFFKSVKREDEHYKIQVKSISGMNVTIIVOSGNLDVESP EGVAASRPTEVSGKLVHANFEVKDFEELSIVNGSLIYVAGEITPAEKVANASF NALGVILDMKNKEPVVEDIALFGHAHGTGDPTPGPPSPNHOPPPSOSSGLPNI PVDTISRAAEKILFGMGESCPARNRHIDSCMLESONWKLIYKNYLKRRLINT GVLTGIEEDPRIVVGAORDALGAVNAASVGTLGLKLNQVSDMTSKGFERSRS IIIASWTAGDFGAVGATEWELEGVLSLHLKATPIINDLVLGTSNFVSASPPLYTL MGINIDVHKPVDGKSLYSDSNWSIKVEELSFDDNAAYEFLLASGLPVSFCFCEADY PYGSTRIDYEALTOKPOLDNOMVRTARAENVAGOLIKLTHDELVDYENTNSKLST MKLNLEKTDIRDMGLSLQMLYSARGDYVSRSLITDPFHAKNEKTREVMREINDRIIM KVYEHPLSPVPSRESPEPHIFEWGSHTLSALVENLKLRQNITAENETLFRNDLAL ATWTGIVANAALSGDIWNIDNEP"			
BASE COUNT	685 a	439 c	515 g 653 t	
ORIGIN				
Query Match	7.5%	Score 215:	DB 12: Length 2292:	
Best Local Similarity	52.1%:	Pred. No. 4, 6e-26:		
Matches 764: Conservative	0:	Mismatches 635:	Indels 68: Gaps 10:	
Oy	974	aggcagcgattatagagcattgacacttggaagtggaaagccctcacacactgcttc	1033	
Dd	887	ACCTTGACACTCTTTGGACAGTCCTCATCTAGGAAGCTGGTGAATCCATCACACCTGGCTTC	946	
Oy	1034	cttccttcaatcaaacaccagttccctccctcagttgcattcatcataagaagcctccacagatcccg	1093	

Db	947	CTTCTTCAATATACATCACTGTTCCGCCAATCTCAGTCAATCAGGCTGCTTAATATACCTG	1006
Qy	1094	cccagcccaatcagtcagacacatgctcccgctgctgaggaagctcaagg-ccctgtg	1152
Db	1007	TGCMAAACAATCTCAAGAGCTGCTGCAGAAAAGCTATTGTGAAAAATGAGAGCAAGCTGTC	1066
Qy	1153	gcccccaagaatgagcaggagagccctcctaggtcccttatcacttggccccggcca	1212
Db	1067	CTGCTATGATGGAACATAGATTCTTCACTGTAAGCTGGAACCTTCA-----CAGAAATCAA	1119
Qy	1213	cgactgcygctagtggttcaacaatcagagacctccaccaccaacacatcttcggc	1272
Db	1120	AATGTAAAGCTCACTATGTGAAAAACGTAAGAAAGAAAGAAATACTTAAACATCTTGGAA	1179
Qy	1273	tgcataagagcgccctcagaagcagatacactagttgtatcaggggcccagagagatgca	1332
Db	1180	GTTATTAAAGGTTATAGAGAACCCAGCCGTATGTGTAGTAGAGACCCAGAGAGAGCT	1239
Qy	1333	tggggcccgag---agcagataatccgctgtgagagagagatactcctgagctgtg	1389
Db	1240	TTCGGGCTGCTGGGTTCGGCGAAGTCCAGTGTGGAAAGGCTTCTGTGTAACCTTGGC	1299
Qy	1390	cgagaccttctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc	1446
Db	1300	CAAGTATCTCAGATATGATTTCAAAAGATGATTTAGACCCAGAGATATATCTTT	1359
Qy	1447	atcagctggaagcggtgtgacttggaaagcggtgctcagagagtgctagaagctac	1506
Db	1360	GCCAGCTGAGTGCAGGACCACTTGGAGCTGTGGTGGCCACTGAGTGTGGAGGAGTAC	1419
Qy	1507	ctcagcgctgctgacactcaagccgtagtgcgtgagagctgagcaaccagagctggg	1566
Db	1420	CTTTCACTCTTGCATTTAAAGCTTTCACCTATATTAATTTGATTAAGTTGCTTGGT	1479
Qy	1567	galtgcaagtttcaatgccaagaccacccctctgacaagttcattagagtgctcg	1626
Db	1480	ACTAGTAACCTTCAAAAGTTCTGCGACCCCTTATATATACATTAATGGAAGATTAAG	1539
Qy	1627	aagcaggtgagttctcccaaccagtgagcagacctctatgaaagagtggtgtccac	1686
Db	1540	CAGAGATGTAAA-----GCATCAGATGATGAAAAATCTATATATAG-----ACAC	1587
Qy	1687	aatccagctgagatgctagtgatcgcgcccctaccatgagcagtgctcctacc	1746
Db	1588	AATGTGATATGCAAAATGTGAAAA-----CTTCTTTCAGCAATGCTGATATCTT	1638
Qy	1747	ttaacggccttgtgagagtcctcgcgtgcagttctccttatggagagcagcagcc	1806
Db	1639	TTCCTTGCATATTTGGAATCCGACAGTTCTTTTGTGTTGTGAGGATG---CAGAC	1695
Qy	1807	taccatctctgcacacaagagagacacttatgaaacctgacataagtgctgcagagc	1866
Db	1696	TATCTTATTTGGGCACTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG	1752
Qy	1867	cgccctgcccgcgtggcccagcgctgcccagctcagagggagagctcctccagctc	1926
Db	1753	CAGCTCAACCAATGTTGTGTACAGCAGGAGAGTGGCTGTGTCATTTATTAACCTT	1812
Qy	1927	agccagatcgccctgctccctcgcactcgcgcgtcagggagagcgctgctccagcagc	1986
Db	1813	ACCCATGAGCTGATTTGAACCTGAGCTATGAGATGTATTAACGCAACATAGCTATTT	1872
Qy	1987	atccggaaacctcaagagttctctgggagactcaagcccgcggctgacctgcagtg	2046
Db	1873	ATGAAGATCTGAAACAGTTTCAAAACAGATATCAGGATATGGCTTAACTACAGTG	1932
Qy	2047	gtgtactgagcgcgaggagactacatcagcgagcgagaaagctgagagagatctac	2106
Db	1933	CTGTATTTCCGCTCGTGAAGACTTCTCGTGTACTTCTAGACATAACAAGTATTTTCA	1992
Qy	2107	agctcggaagagagagcagcagctgacacgcatgtacaagctgagcgaatgctggtg	2166

Db	1993	AATGCTGAGAAAACAACAGATTTTGTATGAGGGAATCAATGATGATTAATGAAGTG	2052
Qy	2167	gagttctactctcttccagtaagtgctgcagccagctcccgcttccgacatctc	2226
Db	2053	GAGTATCACTTCTCTCCCTATGATATCTCCAAAGAGTCTCTTCCGATATCTTC	2112
Qy	2227	atggcgctgagacacagagctgagcccgctgagcagcactgagctgtggtcc	2286
Db	2113	TGGGCTCTGCTCTCAGCTCTCTCAGCTTTAGTGGGAACCTTGAAGCTTGTCAAAA	2172
Qy	2287	aacagctccggagaccgcccgggccactctcctcactgagcttccagagagccgcttc	2346
Db	2173	AATATT-----ACTGCTTTAATGAACCTCTTCA	2205
Qy	2347	cgctgagcctcctgctcactctgagcgtgcagagggagcgaatgctgagagat	2406
Db	2206	AACAGTTGGCCCTGGCTACTGTGACTATTCAGGAGATGCAAAAGCCCTCTGTGTAC	2265
Qy	2407	gtctggaacattgataacaactctga	2433
Db	2266	ATTGGAATATTGACATGATTTTAA	2292

RESULT	10
LOCUS	123809
DEFINITION	Sequence 27 from patent US 5538866.
ACCESSION	123809
NID	91603679
VERSION	123809.1
KEYWORDS	GI:1603679
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	1 (bases 1 to 780)
AUTHORS	Israeli, R.S., Heston, W.D.W. and Fair, W.R.
TITLE	Prostate-specific membrane antigen
JOURNAL	Patent: US 5538866-A 27 23-JUL-1996;
FEATURES	Location/Qualifiers
BASE COUNT	197 a 182 c 201 g 176 t 24 others
ORIGIN	

Query Match 5.6%; Score 160.2; DB 5; Length 780;
 Best Local Similarity 57.1%; Pred. No. 3.9e-17;
 Matches 391; Conservative 0; Mismatches 265; Indels 29; Gaps 5;

Qy	988	ggacatgacagctgagagactgagagaccctacacacagctgctcctcctcaatcaa	1047
Db	17	GGACATGGCCACCTTGGAGCTGAGACCCCTTACACCCAGAGCTTCCCTTCAACAC	76
Qy	1048	accga-cttccctcagttgacatcagcagcttccagcagcagccagccac	1104
Db	77	ACCCANNNGTTTCCACAGTTGATCTTCAAGACATACCCACATGCTTCAACATC	136
Qy	1105	agtcagacattgctcccgctgctgagagagctcaag--ggcctgtggcccccaag	1162
Db	137	TCTACAGTGCAGCAGCCAGGCTGTTCAGCAAAATGATGAGACATGCTGTGAMGN	196
Qy	1163	aatgacagggagcctctagctccttatacactg-----ggc	1203
Db	197	NGTTGAAAGGTGCATCCANNNTCTGTAAAGTTNNACNNNAACAAGAGAGANNNN	256
Qy	1204	cccggcagcagctgagctagtggttcaacaatcagagacctccaccacacac	1263
Db	257	GCCAGANNTAATGTTAAATAGATGTGAATTTCCATGAAGAAGACAGAAATTTGTGAC	316
Qy	1264	atcttcgctgcatcaggaagccgctcagagccagatcactacgtgtgacggggccag	1323
Db	317	ATCTCGTGATATCAGAGGATTTGAAGACCTGATCGGTATGTGTGATTTGAGCCAG	376

ACCESSION AJ012371
NID 94539526
VERSION AJ012371.1 GI:4539526
KEYWORDS naaladase L gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2320)
Pangalos,M.N., Neefs,J.M., Somers,M., Verhasselt,P., Bekkers,M.,
van der Helm,L., Fraiponts,E., Ashton,D. and Gordon,R.D.
Isolation and expression of novel human glutamate carboxypeptidases
with N-acetylated alpha-linked acidic dipeptidase and dipeptidyl
peptidase IV activity
J. Biol. Chem. 274 (13), 8470-8483 (1999)
JOURNAL MEDLINE
REFERENCE 2 (bases 1 to 2320)
Pangalos,M.N.
AUTHORS Pangalos,M.N.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) Pangalos M.N., Biotechnology, Janssen
Pharmaceutica, 30 Turnhoutseweg, 2340 Beerse, BELGIUM
FEATURES
source
1. .2320
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q12"
/tissue_type="small intestine"
17. .2239
/gene="naaladase L"
17. .2239
/gene="naaladase L"
/function="peptidase"
/evidence="experimental"
/product="NAALADase L Protein"
/protein_id="CA39968.1"
/db_xref="PID:e1420518"
/db_xref="PID:94539527"
/db_xref="GI:4539527"
/translation="MONTKYLGLGLAALGLGILLGHPALPKKANSIAPODLLEI
LETVMGOLAHRIENLRELSREPLASSPRDDLYOLLRKDESGDSAEATY
EVLFSPOBPQVNDVPTGSIHSCHREEMVYDGGDPVQPYAAYASCTPO
GLUYANRBAEDPFKLOTOGILECTIALTRIGYGRKAKVNAKRGVAGVLTVD
PAINDGLSPDETFPNSWLPSPGVERKSYIEFGDPLTPYAPVSSFRVLANVS
GEPPIPTPIGQDARDLCLNGLTAPMTWQALCHYRLGFGFDDFPDSCVN
VSYVNRLELRNSNVGLIRGAVEPDRVLYGNHRDSWVGAVDPSSGAVLELSRV
LGLTKGTMPRPSRIVFAWGAEEFGISTEFTFEFNKLOERTVAVIYNDISVEA
NATLRVQGTTPVQSVESATKEIRSPGPDLSIYDMWIRFNSSPVYGLVPSLSLG
AGSDYAPFVHFLGISMDIATYTDKSKSRITPTIHAFDITFDYDKFLDPFSHO
AAVARTGSVILRLSDSFLPLKVSYSFETLRSLQAQODLGLLBQHSISLQPLVTA
VEKFEAAALGORISITLQKSPDPLQVRMLNDQLLETLRNPRAFPEERYSHVL
WAFSHGLRSHIIPGLSNACSRARDTASGSEMAVYQRLSIVTALBGAATLPRVADL
"

BASE COUNT 453 a 725 c 671 g 470 t 1 others
ORIGIN

Query Match 3.2%; Score 91.6; DB 10; Length 2320;
Best Local Similarity 46.6%; Pred. No. 3.7e-06;
Matches 672; Conservative 0; Mismatches 674; Indels 96; Gaps 8;

QY 672 tgaagcgagaaaggtcggagagcagctgcgctgaggaaccctgaactctactgccccta 731
DB 415 tgaaggaanaagctgacccggagagcaaggggccagatgtgtacaaacctatgctgcta 474
QY 732 cagcgcatcggaagctcaaggaagcgtggtgtaagccacacgaaggcgcccgaga 791
DB 475 tgcctcttcttggaaccccaacagggcctctctgtatgccaacggggccggagaga 534
QY 792 cctgcagagacctgcggcgcaagcgtggtacacagctgagccgctgctgctgagcgt 851
DB 535 ctttaagagactacagctcagggcctcaaaccttgaaaggcacacttgcctcgaactcata 594

QY 852 gggggtgatcagcttcgcccagaaggtgaccaatgctcagacttcggggtcaagagct 911
DB 595 tgggggtgtagggcgtggggcccaagcctgtgaacgctgccacacaggggttacctgggg 654
QY 912 gctcatataccagagccagcagactctccaggagcccaagcccaagcctgctccag 971
DB 655 gctggtgtacacagacccttgcgcacatcaacgattgggctgagctcacccagaaactt 714
QY 972 ccagcagcagctga-----tgacatgtgaccttggaacttggaagc 1014
DB 715 tcccaactcctgtactctgccccctcagagatggagagcagggctctactacagagatttt 774
QY 1015 ccccaacacacttgctccctcctcctcaatcaaaccaagcttcctccatgctatc----- 1071
DB 775 tggggaccccttgcacttccctacacttccagccgtcccttcttcccttgcgctggacc 834
QY 1071 -----atcaggtctccagcatcccaagcccatcagtgcaagacttcctcccg 1124
DB 835 caattgtctccggatttccccaattctcttacacagccattgcttccagatccaaagaga 894
QY 1125 cctgcctgaagaaagctcaaaagcctgtggtgcccccaagaattggcagagagcctcctag 1184
DB 895 cctgctctgtatcctcaaggaacttggccccaagcacttgacag-----gaccactggg 951
QY 1185 ctcccctatcaccttggtgcccgggccaagac----- 1217
DB 952 ctccactacagaggtgtgggtccgcttccgcttgacagagacttccacagacagacca 1011
QY 1217 --tgcggtctagtgtaacaacatcacaggaactcaccaccataaacaactctggctg 1274
DB 1012 ggtgaatgtagcgtctacaccccttgagctgaggaactcttccaaactcttgggcat 1071
QY 1275 catcgaagggcgcgtcagagccagatcactagctgtgctatcgggcccaagagggatgat 1334
DB 1072 catccggggggcgtgtgagacctatgcgtacgctgactgtaggaacacagcagacactg 1331
QY 1335 gggcccaaggaagcctaatacgcgtgtggagagcgtatactcttgagactgtgtggagc 1394
DB 1132 ggtgacaggggctgtgacccacagacagtgacacggcgcgtctctgtgagcttccctgt 1191
QY 1395 ctcttcctcatgtgtgagcaagc-----cttcggccccgagagaagctcctctcatag 1451
DB 1192 cctggggacccctctgaagaaggacactgtgctctcgcagatcaactctgtttggcag 1251
QY 1452 ctggaagcgtgtggaacttggaagcgtgtgagcctcacaagatggcgtagaagcctacac 1511
DB 1252 ctggggggcgtgagagagtttggtgctcattggctcaccaggaatttcacagaagaagtctt 1311
QY 1512 cgtgtcgcacctcaaaagcgtagtgtaagcctggaacagcagtgctgtggagatga 1571
DB 1312 caagcttcagagacgcgcagcgtggtcctacatcaacgttggaacatctcggttcccaagc 1371
QY 1572 caagttcatgccaagaccgccccctcttgcaaaagtctcatgtgagagtgctctgaagca 1631
DB 1372 tacccttaggggtcagaggagcgcctccctgtccagagcgtgcttctctcaaaccaaga 1431
QY 1632 ggtggaattcccaacacacagtgtaggaactcctatgaacaggtgtgttcaacaaatcc 1691
DB 1432 gatccgctacacagaccccttgccgacacgtacatctacgaacactgacatccggaacttca 1491
QY 1692 cagctggatgtcgtgaagtgtatccggcccccaaccatga-----cagcag 1736
DB 1492 cccgacagaccccggtgacccctgtggccccaactgttggttctgtgggtcctggcacagca 1551
QY 1737 tgcctatccttcaagcctcttggtgagatccctgcgttgaaattcctcttaagagga 1796
DB 1552 ctatgacaccttgccttcttccctggacatctctccatgacattgcttactatgata 1611
QY 1797 c-----gaccaggcctaacccatcctcgtgacacaaagagaaactta 1841
DB 1612 ccggagcacaactctcagaccagagatctacccacttaccacacagccttcttgcacactttga 1671

QY 1842 gaacctgataaagtgtcgaagccgacctgcccgcgtggtcccaagccgtgaccagct 1901
 DB 1672 CTATGTGACACAGTTTGTGACCCGGCTT---CAGCAGCATCAGGCTGTGCCCCGGAC 1728
 QY 1902 cgcagagcagctccctccatccgctcagcagcagatcgctgtgccccctgacttcgccc 1961
 DB 1729 AGCGGGAGAGTGTGATCTCCGGCTCAGTACAGCTTCTTCGCCCCCAAGTCAAGTGA 1788
 QY 1962 ctacggggagcgtcgtctcagcagacatcgggaaacctcaagagttctctgggacctcaa 2021
 DB 1789 CTACAGTGAACACTCCGAGCTTCTCTGAGGAGCCAGCAAGATCTTGGGGCCTGCT 1848
 QY 2022 gg 2023
 DB 1849 GG 1850

RESULT 15
 LOCUS 123810 660 bp DNA PAT 07-OCT-1996
 DEFINITION Sequence 28 from patent US 5538866.
 ACCESSION 123810
 NID 91603680
 VERSION 123810.1 GI:1603680
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 660)
 AUTHORS Istitelli,R.S, Heston,W.D.W. and Fair,W.R.
 TITLE Prostate-specific membrane antigen
 JOURNAL Patent: US 5538866-A 28 23-JUL-1996;
 FEATURES
 source 1.660 /organism="unknown"
 BASE COUNT 190 a 126 c 147 g 181 t 16 others
 ORIGIN

Query Match 3.1%; Score 88.8; DB 5; Length 660;
 Best Local Similarity 56.0%; Pred.No.1.4e-05;
 Matches 241; Conservative 0; Mismatches 180; Indels 9; Gaps 4;

QY 1217 tgcgctagtggtcaacaatccagagacctccaccccatcaacaacatcttcgggtgca 1276
 DB 96 TGAAGTCACTGTGAACATGTACTGAAGAACAAGAAATACTTAACATCTTTGGCGTTA 155
 QY 1277 tgcgaagccgctcagagccagatcactcgtgtcactcgtgggcccagagagtgatggg 1336
 DB 156 TTAAGGCTATGAGGACACCGCTCATGTGTAGTAGAGCCAGAGAGCGTTGGG 215
 QY 1337 gcccaggg-agcagctaaatccgctgtgggagc-ggctatactcgtgagcgtgtgcgagc 1394
 DB 216 GCCCTGGTNGTCCGAAGTCCAGTGTGGAGACAGGCTTCTGTGAACCTTGCCCAAGT 275
 QY 1395 ctttcctccatggtgagcaac---ggcttcggccccgcagagagctcctcttcacag 1451
 DB 276 ATTCTCAGATATGATTTCAAAAGATGATTTAGACCACAGAGATATATCTTTGCGCAG 335
 QY 1452 ctggagcaggtgggtgacttggagcgtggcctccagcaggtgctagaaagctaccacag 1511
 DB 336 CTGAGCTGACAGAGACTATGAGAGCTGTGTCCGACTGAGTGAGGGGTACCTTTC 395
 QY 1512 cgtgtcgcaccccaagccg----tagtgtacgtgagcctgagacaacgagtcgtgggg 1567
 DB 396 ATCTTTCATCTAAAGNNNGCTTTCATACATTAATNCTGATAAAGTCGTCCTGGGTA 455
 QY 1568 atgacaagttcatgccaaagcccccctctgacaagttcatgtgagagtgctcga 1627
 DB 456 CTAGCAACTTCAAGGTTTCTGCCAGCCCTTATATATACACTTATGGGAAGATATATGC 515
 QY 1628 agcaggtgga 1637

DB 516 AGANNNGTA 525

Search completed: January 13, 2000, 02:55:32
 Job time: 2290 sec

This Page Blank (usp10)

1	148.2	5.2	779	065535	Sequence used in
2	127	4.4	540	065537	Sequence used in
3	89.4	3.1	93	T19431	Human gene signature
4	88.8	3.1	660	065536	Sequence used in
5	53.8	1.9	2653	065520	Prostate-specific
6	53.8	1.9	2654	T36785	Prostate-specific
7	46.2	1.6	1030	V99230	DNA encoding an
8	46	1.6	49272	V35000	DNA encoding an
9	45.8	1.6	53789	V21187	Mycobacteriophage
10	45	1.6	4356	Q37543	Amphicollapsin med
11	45	1.6	4356	Q95540	Cardiac adenylyl
12	44.8	1.6	5821	T58686	DNA encoding S. fi
13	44.8	1.6	2135	X26879	DNA encoding a pro
14	44.6	1.6	1164	V32494	S. griseofuscus g
15	44.4	1.5	26338	V62134	HSV-2 strain SB5
16	44.4	1.5	20387	V62159	HSV-2 strain SB5
17	44.4	1.5	117213	V62126	HSV-2 strain SB5
18	44.4	1.5	114955	X53491	Human adenovine A
19	44	1.5	1389	T18534	Chimeric heretigul
20	44	1.5	5676	V21186	Amphicollapsin med
21	44	1.5	53789	V21187	Amphicollapsin med
22	43.8	1.5	2846	V41260	Chlamydomonas reli
23	43.2	1.5	44377	T78508	Platenolide synth
24	43.2	1.5	44377	T80414	Platenolide synth
25	43	1.5	1278	Q21629	Bleomycin acetyl
26	42.8	1.5	15079	Q91360	S. clavuligerus c
27	42.8	1.5	8065	V38335	Manic-depressive
28	42.8	1.5	6093	X25775	S. erythraea oleana
29	42.6	1.5	1542	X16151	Human Sox1 encodi
30	42	1.5	2729	N92629	DNA fragment cont
31	42	1.5	1164	N92631	DNA fragment cont
32	42	1.5	1630	Q66684	Acyl B2 gene - enc
33	42	1.5	12001	Q76213	HSV L/ST region. I
34	42	1.5	4257	V10362	Infected cell prot
35	42	1.5	4257	V68520	The nucleotide seq
36	41.8	1.5	15875	T68715	Streptomyces vene
37	41.6	1.4	2072	Q04529	Elastase gene. Pst
38	41.6	1.4	3059	N70566	Sequence encoding
39	41.6	1.4	2483	T12462	Human K+ channel

QY	988	ggaacttgcaccttgggaacttggagaagcccttaacacacttggcttccctcttcataca	104
Db	17	GGACATCCCAACCTTGGAACTGGAGACCTTACACCCAGC--TCCCTTGTTCAACCA	75
QY	1048	accac---gttccttcagttgtcatcatcagagccttcceagcatcccaagcccaacc	110
Db	76	ACCCANNNGTTTCCACAGATTGAATCTTACGAGACTACCCACATTGCTGTTCAGACATC	135
QY	1105	agtgacagacattgcctccgcctctgtcttgaagaactcaag--gcctctgtgcccccaag	116
Db	136	TCTAGCAGTGCAGACGACGACGAGCTGTTCAGCAAAATGGATGGAACACATGCTCTGANA	195
QY	1163	aatgacggggagagctctcagctccacctataactg-----ggc	120
Db	196	NGTTGGAAAGTGTGATCCANNNTTCTCTGAAGCTNNGACNNAACAAGAGAGANNNN	255
QY	1204	cccgagccacagcactgcgcgtcagtggtcaacaatcacagagaccccaaccatcaacaac	126
Db	256	GCCAGATTAATGTGGAACACTAGATGTGACAAATTCCATGAAACACAGGAAGATTTCGAA	315
QY	1264	atcttcgctctgcatcgaaggccgctcagagccagatcaactactgltgcatcggggccca	132
Db	316	ATCTTCGGTCTACTCCAGGGATTTGGAAGACCTGATCGTATGTGTGATTGGAGCCCA	375
QY	1324	aggagatcatcggggcccaagagcagctcaaatccgcctgttggggcggctatactcccgag	138
Db	376	AGAAGTCTCTGGGGCCAGGAGTGCGCTAAAGCTGGGCACTGGAACTCTCTATTATTGGTAA	435

```
OY 1384 ctggtcgagacattctccatggtgagcaac---ggcttcggcccccagaaatctc 1440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 436 CTGGCCCGGTATCTCAACATAGTGAAGGAGGCGTTAAACCGGCGAAGCATC 495
OY 1441 cttctcatcagctggagcgggtgtgactttgaaagcggtggctccaggaagtgtagaa 1500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 496 AACTTTGGTAGCTGAGTCAAGAGACTACGAGCTGTGGGTGTACTAATAGCTGGAG 555
OY 1501 gctacccatcagctgtcgtgacactcaagcgcgtagtgtatgta--gcctggacaacgag 1558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 556 GGGTACTGTGCATGCTGCATGCCAAAGCTTCACTTACATCANNCGTATGCTCCAG 615
OY 1559 tgcctggggatgacaaattcctgccaagacagccccccttcgacaagctcatgaga 1618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 616 TCCTGGAGCAAGCCATGTCAAGATTTCGCCAGCCCTTGTGTATGCTGTGGGA 675
OY 1619 gtgtcctgaagcaggtggaattctc 1643
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 676 GTATTATGAAGGGGTGAAGAAATCC 700

RESULT 2
ID 065537 standard; cDNA: 540 BP.
AC 065537:
DE 11-JAN-1995 (first entry)
DT Sequence used in identification of PSM Ag cDNA.
DE Prostate-specific membrane antigen; PSM: prostate cancer;
KW transmembrane glycoprotein; imaging; targeting; tumour detection;
KW antibody detection; ds.
OS Homo sapiens.
PN WO9409820-A.
PD 11-MAY-1994.
PE 05-NOV-1993; U10624.
PR 05-NOV-1992; US-973337.
PA (SLOK) SLOAN KETTERING INST CANCER.
PI Fair WR, Heston MDW, Israeli RS;
DR WPI: 94-161729/20.
PT Prostate-specific membrane antigen and DNA encoding it - is
PT useful for detecting haematogenous micro-metastatic tumour cells
PT and for identifying ligands which bind to PSM Ag
PS Disclosure: Page 121; 196pp; English
CC This sequence is given in the specification but is not mentioned in
CC the disclosure. The PSM coding sequence is useful for suppressing or
CC modulating the metastatic ability of prostate tumour cells to grow,
CC or for eliminating them. The protein is useful to identify or purify
CC ligands of the Ag. It is also an attractive target for Ab-directed
CC imaging and targeting of prostatic tumour deposits.
SQ Sequence 540 BP; 152 A; 106 C; 133 G; 147 T;

Query Match 4.4%; Score 127; DB 1; Length 540;
Best Local Similarity 60.4%; Pred. No. 1.5e-17;
Matches 261; Conservative 0; Mismatches 166; Indels 5; Gaps 3;

OY 1217 tgcgctagtgtcaacaacacagactccaccctcaacaacatcttcgctgca 1276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 75 TCAAGCTACTGTGACCAATGTGCTGAAGATAAAATTTCTTAACATCTTGGAGTTA 134
OY 1277 tcgaagcgccctcagacagatcaactacgtgtcaltcagggcccccaggaagtgcag 1336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133 TTAAGGCTTTGTAGAACCAAGATCACTATGTGTAGTTGGGCCCAAGAGATGATGG 194
OY 1337 gcccaagagcaagtaaat-cgctgtgggagcgctatactcctgagctgtgagacc 1395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 195 GCCCTGGACCTCCAATAATNCGCTGTAGGCACAGCTCTCTATTGAAACTGCCAGATG 254
OY 1396 ttcttcctcatgg---ttagcaagcgcttcggcccccgaagaagctcctctcatcagc 1452
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 255 TTCTAGATATAGTCTTAAGAGATGGGTTTCAGCCAGCAAGCAATTAATTTGGCACT 314
OY 1453 tggagcgtgtgtgactttgaaagcgtgggtccacgagatgagctgaagcctacc-tcag 1511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
DB 315 TGAAGTGTGAGAGACTTTTGATCGGTGTGTGTCACATGAATGGCTAGAGGATTAOCTTCG 374
OY 1512 cgtgtgcacctcaagccgtagtgcagtgagcctgagacaacgagtgctggggagtga 1571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 375 TCNCTGCTATTTAAAGGCTTTCACCTATATTAACCTGATTAAGGCGGTCTTGTGACAG 434
OY 1572 caagttcatcccaagaccacccccctctcgaagtcctcttgaagtgagtgctcgaagca 1631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 435 CAACTTCAGGTTTCTGTCGACGCCACCTGTGTATACGCTTATGAGAAACAAATGCANAA 494
OY 1632 gttgagattctc 1643
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 495 TGTGAAGCATCC 506

RESULT 3
ID T19421 standard; cDNA to mRNA; 93 BP.
AC T19421:
DE 04-JUL-1996 (first entry)
DE Human gene signature HUMGS00446.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PE 11-JUN-1995.
PR 12-NOV-1994; J01916.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI: 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function. By preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 380; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (gene signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 93 BP; 28 A; 17 C; 24 G; 23 T;

Query Match 3.1%; Score 89.4; DB 1; Length 93;
Best Local Similarity 97.8%; Pred. No. 5.1e-10;
Matches 90; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2779 gatcgctggccacatagcctatgagccaaagcgtgtgtgtggaagggcgctgggt 2838
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 GATCGTTGGCACACATAGCCTTATGGCCAAAGGTGTGTGTGTAAGGGGCGGTGAGT 60
OY 2839 ttcaatatcaataaaccacccgatatacaataa 2870
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 TTCAATATCAATTAACCAACCATGATATCAATAA 92

RESULT 4
ID 065536 standard; cDNA: 660 BP.
AC 065536:
```

DT 11-JAN-1995 (first entry)
 DE Sequence used in identification of PSM Ag cDNA.
 KW Prostate-specific membrane antigen; PSM; prostate cancer;
 KW transmembrane glycoprotein; imaging; targeting; tumour detection;
 KW antibody detection; ds.
 OS Homo sapiens.
 PN WO9409820-A.
 PD 11-MAY-1994.
 PE 05-NOV-1993: US-973337.
 PR 05-NOV-1992: US-973337.
 PA (SLOK) SLOAN KETTERING INST CANCER.
 PI Fair WR, Heston WDM, Israeli RS;
 DR P-PSDB: R55097.
 PT Prostate-specific membrane antigen and DNA encoding it - is
 PT useful for detecting haematogenous micro-metastatic tumour cells
 PT and for identifying ligands which bind to PSM Ag
 PS Disclosure; Page 120; 196pp; English
 CC This sequence is given in the specification but is not mentioned in
 CC the disclosure. The PSM coding sequence is useful for suppressing or
 CC modulating the metastatic ability of prostate tumour cells to grow,
 CC or for eliminating them. The protein is useful to identify or purify
 CC ligands of the Ag. It is also an attractive target for Ab-directed
 CC imaging and targeting of prostatic tumour deposits.
 SQ Sequence 660 BP; 190 A; 126 C; 147 G; 181 T;

Query Match 3.1%; Score 88.8; DB 1: Length 660;
 Best Local Similarity 56.0%; Pred. No. 9.3e-10;
 Matches 241; Conservative 0; Mismatches 180; Indels 9; Gaps 4;

QY 1217 tgcgcagtagtgatcaacatccagagaccctcccccacacacatcttcggcgca 1276
 DB 96 TGAAGCTACGTGATGACATGACTGAAAGAAACAGAACTTAACTTTGGCGTTA 155
 QY 1277 tgcgaagccgcgtcagagccagatcactacgttcatcggggcccaagagagatcgatgg 1336
 DB 156 TTAAGGCTATGAGAGACAGACCGCTACATGTGTAGAGAGCCAGAGACGCTTGGG 215
 QY 1337 gcccagag-agcagctaaatccgcgtgtggagac-ggctactcctcgagagctggtcgagac 1394
 DB 216 GCCCTGGTNGTGGCAAGCTCAGTGTGGAGACGCTTTCGTTGGAACCTTGCCCAAGT 275
 QY 1395 ctttcccccactggtgagcaaac---ggcttcggcccccagagaatcccttcctacatag 1451
 DB 276 ATTCTCAATATGATTTTAAAGATGATGATTTAGACCCAGACAGATATTAATTTGCCAG 335
 QY 1452 ctggagacggtgtgacttggaaagcgtgtggtccagagagtgctagagagctaccatag 1511
 DB 336 CTGGACTGCAGAGAGACTATGGAGCTGTGTGCTCGACTGAGTGGGAGGGGTACTCTTC 395
 QY 1512 cgtgctgacctcaaacgcg----tagtgcgtgagccttgacacacgcaatgctggggg 1567
 DB 396 ATCTTTCATCTAAAGNNNGCTTTCACTTACATTAATCTGATTAAGTTCGCTCGGATA 455
 QY 1568 atgcacaatttcacatccaaagccagcccttcgtacacagttctattagagtgccgga 1627
 DB 456 CTAGCAACTTAAAGGTTTCTGCGACGCCCTATTATTAATCACTTATGGGAGAGATAATGC 515
 QY 1628 agcaggttga 1637
 DB 516 AGGANNCTA 525

RESULT 5
 ID Q65520 standard; cDNA; 2653 BP.
 AC Q65520:
 DT 11-JAN-1995 (first entry)
 DE Prostate-specific membrane antigen cDNA.
 KW Prostate-specific membrane antigen; PSM; prostate cancer;
 KW transmembrane glycoprotein; imaging; targeting; tumour detection;
 KW antibody detection; ds.
 OS Homo sapiens.

FH Key Location/Qualifiers
 FT cds 262..2514
 FT /*tag= a
 FT /product= prostate specific membrane antigen (PSM)
 PN WO9409820-A.
 PD 11-MAY-1994.
 PE 05-NOV-1993: US-973337.
 PR 05-NOV-1992: US-973337.
 PA (SLOK) SLOAN KETTERING INST CANCER.
 PI Fair WR, Heston WDM, Israeli RS;
 DR P-PSDB: R55097.
 PT Prostate-specific membrane antigen and DNA encoding it - is
 PT useful for detecting haematogenous micro-metastatic tumour cells
 PT and for identifying ligands which bind to PSM Ag
 PS Claim 3; Page 103-106; 196pp; English
 CC Q65520 encodes a prostate specific membrane antigen (PSM, R55097).
 CC The PSM coding sequence is useful for suppressing or modulating the
 CC metastatic ability of prostate tumour cells to grow, or for
 CC eliminating them. The protein is useful to identify or purify ligands
 CC of the Ag. It is also an attractive target for Ab-directed imaging
 CC and targeting of prostatic tumour deposits.
 SQ Sequence 2653 BP; 782 A; 524 C; 640 G; 707 T;

Query Match 1.9%; Score 53.8; DB 1: Length 2653;
 Best Local Similarity 48.8%; Pred. No. 0.016;
 Matches 145; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 1255 atcaacacatcttcgcgtgcatcgaagcgccgcacagagacatcactacgtgtc 1314
 DB 1324 ATTATCAATGTGATNGTACTCTCAGAGAGACAGTGGACACAGACATATGCTATTGTG 1383
 QY 1315 ggggcccagagatgcatgagggcccaagagcaatccgcgtgtggtggagcgctata 1374
 DB 1384 GGAGGTACCGGAGACTCATGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1443
 QY 1375 ctctcgagagctggtgagcacttcttcctcatgtgtgagcaagcgttcggcccgaga 1434
 DB 1444 GTTCATGAATGTGAGAGAGCTTTGGAACACTGAAAAAGGAAGGTGGAGACTGAAGA 1503
 QY 1435 agtctccctcatcagctgagcggtgtgacttggaaagcgtgtggtccacagagtg 1494
 DB 1504 ACAATTTGTTTGCAGCTGGGATGACAGAAATTTGCTCTTCTTCTTCTTCTTCTTCTTCT 1563
 QY 1495 ctagaagctacccacgtgtgctgacacctcaaacgcgtagtgatcgtgagcctggag 1551
 DB 1564 GCAGAGGAGAAATTCAGAGCTCTTCAAGAGCGTGGCGCTTATTAATGATGCTGAC 1620

RESULT 6
 ID T36785 standard; cDNA; 2654 BP.
 AC T36785:
 DT 04-NOV-1996 (first entry)
 DE Prostate-specific membrane antigen cDNA.
 KW Prostate-specific membrane antigen; PSM; promoter; prostate cancer;
 KW metastasis; gene therapy; diagnosis; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT 5'utr 1..261
 FT /*tag= a
 FT /*tag= b
 FT cds 262..2253
 FT /product= PSM antigen
 FT misc-feature 114..380
 FT /*tag= C
 FT /note= "bases 114-380 (-147 to +109) are absent
 FT in PSM" cDNA"
 FT 2352..2357
 FT /*tag= d
 PN WO9626272-A1.
 PD 29-AUG-1996.

PF 23-FEB-1996: U02424.
PR 24-FEB-1995: US-394152.
PR 02-JUN-1995: US-470735.
PR 02-JUN-1995: US-466381.
PA (SLOK) SLOAN KETTERING INST CANCER RES.
PI Fair WR, Heston MDW, Israeli RS;
DR WPI: 96-402365/40.
P-PSDB: W02234.
CC DNA encoding alternatively spliced prostate-specific membrane
PT antigen - useful to develop prods. for detecting haemotogenous
PT micrometastatic tumour cells, or prostate cancer progression
PS Example 1: Fig 47A-D: 284pp; English.
CC A cDNA clone (T36785) codes for human 100 kDa prostate-specific
CC membrane (PSM) antigen (W02234), an integral membrane glycoprotein
CC that is very highly expressed in prostatic tumours and metastases.
CC It was obtd. from lymph node carcinoma of prostate (LNCaP) cell
CC mRNA by PCR amplification (see also T36795-808) and screening of an
CC LNCaP cDNA library using an amplified cDNA partial clone as probe.
CC The cDNA can be used to provide probes and primers useful e.g. in
CC detecting haemotogenous micrometastatic tumour cells and determining
CC prostate cancer progression (see also T36827-30 and T36813-18),
CC and in gene therapy. An alternatively spliced PSM, PSM', has a
CC shorter cDNA sequence. PSM genomic DNA is given in T36786.
SQ Sequence 2654 BP; 782 A; 525 C; 639 G; 708 T;

Query Match 1.9%; Score 53.8; DB 1; Length 2654;
Best Local Similarity 48.8%; Pred. No. 0.016;
Matches 145; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 1255 atcaacaacatctcggtcgtcaggaagccgtcagagccagatacactcgtgtcatic 1314
DB 1324 ATTTAAATGTAGTACTCTCAGAGAGCGAGTGAACACAGATATGTCATTCTG 1383
QY 1315 gggggcccaaggagatcgtggtggcccaaggacagctaaatccgctgtggggagcgctata 1374
DB 1384 GGAGGTACCGGACATCATGGGTGTTGGTGTATGACCCCTCAGATGGAGACGTGTT 1443
QY 1375 ctctcgtgagctgtgtcggacatttctccatgtgtgagcaagcgctccggcccgaga 1434
DB 1444 GTTACGTAATGTTGTAGAGAGCTTTGGACACCTGAAAAAGGAGGTTGAGACCTAGAACG 1503
QY 1435 agctcctcttcacgtcggagcggtgtgtgacttgggaagcggtgggtcccaagcgatgg 1494
DB 1504 ACAATTTGTTTTCAGAGCTGGGATGCAGAAAGATTGTTGCTTCTTGTCTTACTGTGCG 1563
QY 1495 ctagaaggtcactcagcgtgtcgtcacctcaagccgtagtgatcgtgagccttgagac 1551
DB 1564 GCAGAGGAGATTCAGAGACTCTTCAAGACCGTGGCTGTATATATGCTGAC 1620

RESULT 7
V99230
ID V99230 standard; DNA: 1030 BP.
AC V99230;
DE 09-MAR-1999 (first entry)
DI DNA encoding an active acyltransferase domain (venAT) from a PKS.
KW Enzymatically active acyltransferase domain; polyketide synthase; PKS;
KW polyketide synthetis; antimicrobial; anticancer; antifungal;
KW immunosuppressant; antihelminthic; venAT; malonate acyltransferase domain;
KW PKS cluster; Streptomyces venezuelae ATCC15439; ds.
OS Streptomyces venezuelae.
PN W0951695-R2.
PD 19-NOV-1998.
PR 13-MAY-1998: U09518.
PR 16-MAY-1997: US-858003.
PA (ABBO) ABBOTT LAB.
PI Kakavas SJ, Katz L, Pereda-Lopez A, Ruan X, Staasi DL,
PI Summers RG;
DR WPI: 99-070114/06.
P-PSDB: W87713.
PT New polyketides produced by microorganism having altered polyketide
PT synthase genes - especially chimeras containing a heterologous

PT acyltransferase domain, useful as e.g. antimicrobials, anticancer
PT agents
PS Claim 24; Fig 13; 137pp; English.
CC The present sequence encodes an enzymatically active acyltransferase
CC domain, venAT, from a polyketide synthase (PKS). The venAT protein
CC is derived from the malonate acyltransferase domain from the
CC PKS cluster of Streptomyces venezuelae ATCC15439. The sequence
CC can be used to replace acyltransferase domains of other PKS enzymes,
CC which are then used to generate novel polyketides by de novo biosynthesis
CC rather than by chemical modification. Polyketides of the invention have
CC the methyl groups on the macrolactone ring substituted with -H, -Et
CC and/or -OH, and the ethyl side chain substituted with hydroxymethyl or
CC dihydroxycyclohexylmethyl side chains. Polyketides of the invention are
CC antimicrobial, anticancer, antifungal, immunosuppressant and/or
CC antihelminthic agents, or their intermediates.
SQ Sequence 1030 BP; 132 A; 383 C; 362 G; 153 T;

Query Match 1.6%; Score 46.2; DB 1; Length 1030;
Best Local Similarity 47.4%; Pred. No. 0.48;
Matches 138; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 694 cagctgcgcgtgaggagacctgactctactgcccctacagcgccatcggaacgtcacg 753
DB 238 CMTTCGGGCTGTTCAGGCGCCAGGCGCTCTGCCGACTACTCTATGCGCCATCATC 297
QY 754 ggaagctgtgtgtacgcccactaagggcgccgaagactgcaagactcgtggccagg 813
DB 298 GCGGAATGACCGCGCGCCACCTGCGCGGGCTCTCGATCTGGCGAGCGCGTCTCTG 357
QY 814 ggcgtgagatccagatggcgccctgtgtcgtgtggtgcgtgggtgtatatagcttccca 873
DB 358 GTGCGCCACCGCGCGCGCTGTATGCAGTGGCGCCGCGCGCGTGTGGCCCGCGTTC 417
QY 874 aaggtgaccatgtcctcagacttcgggtcgaagagtgatcatataccagagccagcg 933
DB 418 CAGGCGAGGAGAGAGAGGTATCGCAGAGCGCTCGGACCTTGAGAGATGGGTTCCGTG 477
QY 934 gactctcccaagaccaccaagccaagcctgtctcagccagccagcgagt 984
DB 478 GCGGAGTCAACGCGCCGACCCACCTGCTCTCCGCGAGAGAGCGCG 528

RESULT 8
V35000
ID V35000 standard; DNA: 49272 BP.
AC V35000;
DE 27-AUG-1998 (first entry)
DI Mycobacteriophage D29 DNA.
KW Shuttle plasmid; vector; mycobacteriophage; lambda phage cosmid; antigen;
KW anti-tumour agent; lymphokine; immune response; tolerance; mutant; toxin;
KW autoimmune disease; antiserum; vaccine; humoral response; T-cell memory;
KW cellular response; mucosal response; diagnostic; antibody;
KW infectious agent; ds.
OS Mycobacteriophage D29.
PN US573267-A.
PD 30-JUN-1998.
PR 07-MAR-1996: 614770.
PR 07-MAR-1996: US-614770.
PR 07-FEB-1992: US-833431.
PR 29-APR-1993: US-057531.
PR 23-MAY-1994: US-247901.
PA (UYPT) UNIV PITTSBURGH.
PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
PI Hatfull GF, Jacobs WR;
DR WPI: 98-387007/33.
PT Nucleic acid encoding the D29 mycobacteriophage and related shuttle
PT plasmids - for producing recombinant mycobacteria and auxotrophic
PT mutants to express antigens and proteins providing long lasting T
PT cell memory.
PS Claim 1; Column 7-50; 28pp; English.
CC This sequence encodes the D29 mycobacteriophage which is used in the
CC construction of a new shuttle plasmid containing D29 with an E. coli

CC lambda phage cosmid inserted in a non-essential region. This plasmid is
 CC used to insert nucleic acid into mycobacteria by (non-)homologous or
 CC site-specific recombination, particularly for expressing antigens,
 CC anti-tumour agents, lymphokines etc., e.g. for evoking an immune response
 CC or to induce tolerance to autoimmune diseases such as rheumatoid
 CC arthritis. Such plasmids that contain a transposon or recombinant DNA can
 CC be used to generate libraries of mutant mycobacteria, e.g. where
 CC expression of recombinant mycobacterial DNA (to produce an antisense
 CC molecule) will inactivate a selected mycobacterial gene. These mutants
 CC are useful for selection of vaccine strains for inducing humoral,
 CC cellular and/or mucosal responses and for expressing antigens for use
 CC as vaccines or diagnostic agents, also other useful proteins such as
 CC enzymes. Also shuttle plasmids specific for a particular Mycobacterium
 CC can be used diagnostically, i.e. they are ligated to a reporter gene,
 CC under control of a strong mycobacterial promoter, then incubated with
 CC test sample and then examining for expression of the reporter. Vaccines
 CC based on the new mutants prime long-lasting T cell memory, stimulating
 CC a secondary antibody response that neutralises infectious agents or
 CC toxins.
 SQ Sequence 49272 BP; 8912 A; 15900 C; 15401 G; 9059 T;

Query Match 1.6%; Score 46; DB 1; Length 49272;
 Best Local Similarity 51.5%; Pred. No. 0.98;
 Matches 106; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 130 gaggaagaagaagaacgagggagggcgagagacattgcccactctgcccacatg 189
 DB 15887 GAGGACGACGAGAGGCGCGGAGCTGTGGAGAGTACTCGCTCGCATTTGCGAGATC 15946
 QY 190 gagctgagggcgccctgagccctgagccttagagccagagcaaacctcctctg 249
 DB 15947 GTCCCAAGGCTCTCCGGCTGATGCGCACCAAGAGTGATCGCGGCTGGAGC 16006
 QY 250 gcgcagcagagagcagggagcgcctaccctgctcctgagccctgctgactcctact 309
 DB 16007 GAGGAAGAGAGATCCAGAGATCCGGCGGACCTGTACCGCACCTTACTGCGAGCTGATG 16066
 QY 310 ggggcttctcactgaggtactgctgc 335
 DB 16067 GTGGAGACGCACTGGGGGAGCCGC 16092

Result 9
 V21187
 ID V21187 standard; DNA: 53789 BP.
 AC V21187;
 DE 24-JUL-1998 (first entry)
 KM Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
 KM Amycolatopsis mediterranei rifamycin synthesis; gene cluster;
 OS Amycolatopsis mediterranei.
 FH Key Location/Qualifiers
 FT 1825..15543
 FT /tag= a
 FT /label= ORF_A
 FT /product= "polyketide synthase"
 FT 15550..30759
 FT /tag= b
 FT /label= ORF_B
 FT /product= "polyketide synthase"
 FT 30893..36060
 FT /tag= c
 FT /label= ORF_C
 FT /product= "polyketide synthase"
 FT 36259..41325
 FT /tag= d
 FT /label= ORF_D
 FT /product= "polyketide synthase"
 FT 41373..51614
 FT /tag= e
 FT /label= ORF_E
 FT /product= "polyketide synthase"

FT CDS 51713..5293
 FT /tag= f
 FT /label= ORF_F
 FT /product= "polyketide synthase"
 PN MO9807868-A1.
 PD 26-FEB-1998.
 PF 18-AUG-1997; E04495.
 PR 20-AUG-1996; EP-810511.
 PA (NOVS) NOVARTIS AG.
 PI Engel N, Schupp T, Toupet C;
 DR WPI: 98-169172/15.
 DR P-PSDB: W52845-W52850.
 PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
 PT to produce rifamycin and rifamycin analogues
 PS Claim 4; Page 53-102; 205pp; English.
 CC The present sequence represents a Amycolatopsis mediterranei rifamycin
 CC synthesis gene cluster DNA fragment from the present invention. The
 CC DNA fragment comprises a DNA region involved directly or indirectly
 CC in the gene cluster responsible for rifamycin synthesis, including
 CC the adjacent DNA regions to the right and left which, by reason of
 CC their function in connection with rifamycin biosynthesis, qualify
 CC as constituents of this rifamycin gene cluster, and functional
 CC fragments, derivatives or constituents of these. The Amycolatopsis
 CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
 CC for producing rifamycin, rifamycin analogues or precursors. It can also
 CC be used for inactivating or modifying genes involved in ansamycin or
 CC rifamycin biosynthesis. The DNA can be used for constructing mutant
 CC actinomycetes strains from which the natural rifamycin or ansamycin
 CC biosynthesis gene cluster has been partly or completely deleted. The
 CC DNA fragment can be used for assembling a library of polyketide
 CC synthases, which can be used for assembling a library of polyketides.
 CC A hybridisation probe of the invention can be used for identifying DNA
 CC fragments involved in the biosynthesis of ansamycins.
 SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T;

Query Match 1.6%; Score 45.8; DB 1; Length 53789;
 Best Local Similarity 44.1%; Pred. No. 1.1;
 Matches 191; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

QY 1877 ccgtggcccaagcgccgtgagccagctgcagggagagcttctcactcggccagccagcagatc 1936
 DB 27737 CCTTGGCCCGCATCGAGGCGCAGGCCCTCCGAGGTCCTTTTACTGTACCCCTGATGGTG 27796
 QY 1937 gcccgtgcgccctgcacttcgcgcgtacagggagcgtgcttccctcaggaacatcgggaacc 1996
 DB 27797 ACTGATCCGCTGACCGCGGGATCGTGCAGCGGCGCTACGTGACCGAAGCTCGCGAAC 27856
 QY 1997 tcaagagttctcttggggaactcaaggccgcgggctgacacctgcagtggtggtactcg 2056
 DB 27857 AGGTCGGGTTCGGTCCGGCGCTCGCGAGACTCTGTCGCCAGGCGCACGGGGTTCGTG 27916
 QY 2057 cgcgggggagactacacccggcgcgcgagaaagctgagcaggaatcctcagctggag 2116
 DB 27917 AGGTCAAGCGCGACCCGGGTGCTGTCCAGCCCTCATGAATCATCAGCAGAGCGGCTG 27976
 QY 2117 agagagacgagcagctgacacgcatgtacaacgtgagcacaatgcyggagttact 2176
 DB 27977 TGACCGGGTGCCTGCGGCGCGAAGAGCGTGGCTGTGCGCGCGCTGACGTGATGGCG 28036
 QY 2177 tccctccagtaagtgtgcgcagccagactcccgcgttcgcgcacatctcattgagcgctg 2236
 DB 28037 AGCTGTACTGTGACGGGTGTCCGCTGACTGACCGCGGTCTCGCGCGAGCGCGG 28096
 QY 2237 gagagacacgctggcgccctgctgtgacacactgsgtgcgtcagcctcaagaagctcg 2296
 DB 28097 TCGACCTGCCGAAGTACGCTTGTGACCGAGGCACTACTGCTGCGCGCGCGAGTCCG 28156
 QY 2297 ggaaccccgaggc 2309
 DB 28157 CGACGACGCGGC 28169

```

RESULT 10
037543/c
ID 037543 standard; cDNA; 4356 BP.
AC 037543;
DT 17-JUN-1993 (first entry)
DE Cardiac adenylyl cyclase type V gene.
KW CACV; therapy; diagnostic; cardiac function; cyclic AMP; cAMP; heart;
OS failure; ss.
FH Canis familiaris.
FT Key Location/Qualifiers
FT cds 148..3702
FT /*tag= a
EP-529622-A.
PN 03-MAR-1993.
PD 27-AUG-1992; 114637.
PR 29-AUG-1991; US-751460.
PA (AMCY ) AMERICAN CYANAMID CO.
PI Ishikawa Y, Koniski AF;
PI WPI; 93-068688/09.
DR P-PSDB; R32882.
PT Isolated nucleic acid mol. encoding Cardiac adenylyl cyclase type
PS V - useful for determining and modifying cardiac function
PS Claim 1; Page 15-27; 38pp; English.
CC Left ventricular tissue of canine heart was used as a source of mRNA.
CC A cDNA library was prepd. in lambda gt10 phage. A 970 bp Aat-HincII
CC fragment from type I adenylyl cyclase cDNA was used as probe. The
CC clones isolated were used to obtain cDNA encoding CACV. This probe
CC may also be used to screen a human cardiac cDNA library to obtain
CC the cDNA encoding human CACV. CACV, its analogues and antibodies
CC are useful in therapy or diagnostic assays, e.g. in modifying and
CC determining cardiac function. A decrease in CACV content of the
CC heart contributes to impaired cAMP produ. and in heart failure. The
CC CACV can also be used to screen for cpds. which stimulate or inhibit
CC the activity of the cyclase.
SQ Sequence 4356 BP; 837 A; 1367 C; 1377 G; 775 T;

Query Match 1.6%; Score 45; DB 1; Length 4356;
Best Local Similarity 49.4%; Pred. No. 1.1;
Matches 117; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

OY 1864 ggcgcgtccgcgcgtggccagcgtgcccagctgcagcgagcagctctcctcag 1923
DB 469 ggcctcggcgctggcgctggcgccggccggccggccgctgctcagctccac 410
OY 1924 ctcaagcagatcgctctgctccctgacttcgagcgtcagggagctgctcctcag 1983
DB 409 cGAGCGGGGGCGCACCTCCGCCCGCCCGCCGCGCGCGCGCGCGCGCGCAAG 350
OY 1984 caccatcggaacctcaagagttctctggaacctcaagggccggcggtgacctgag 2043
DB 349 GCGCGCCCGGAGACTGCCCGCCCGCCCGCCCGCCCGCCCGCGCGCGCGCGCG 290
OY 2044 tgggtgtactcgcgcgggggagactacatccgcgcgcggaagctgcgcagag 2100
DB 289 CAGTGTGCGCGCGCGGCTCTGCCAGCGCCGACCTGGAGCGGAAGTGAAGCGAAG 233

RESULT 11
095540/c
ID 095540 standard; DNA; 4356 BP.
AC 095540;
DT 31-JAN-1996 (first entry)
DE Cardiac adenylyl cyclase gene.
KW Cardiac adenylyl cyclase; effector enzyme; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FH cds 148..3702
FH /*tag= a
TW-243453-A.
PN 21-MAR-1995.
PD 02-JUL-1992; 105242.
PR 12-JUN-1992; US-899068.

```

```

PA (AMCY ) AMERICAN CYANAMID CO.
DR WPI; 95-214006/28.
DR P-PSDB; R78519.
PT Cardiac adenylyl cyclase and corresp. DNA - having specified
PT sequences
PS Claim 1; Fig 2; 45pp; Chinese.
CC 095540 encodes R78519, the novel effector enzyme cardiac adenylyl
CC cyclase.
SQ Sequence 4356 BP; 836 A; 1372 C; 1373 G; 775 T;

Query Match 1.6%; Score 45; DB 1; Length 4356;
Best Local Similarity 49.4%; Pred. No. 1.1;
Matches 117; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

OY 1864 ggcgcgtccgcgcgtggccagcgtgcccagcgtgcccagctgcagcgagcagctcctcag 1923
DB 469 ggcctcggcgctggcgctggcgccggccggccggccgctgctcagctccac 410
OY 1924 ctcaagcagatcgctctgctccctgacttcgagcgtcagggagctgctcctcag 1983
DB 409 cGAGCGGGGGCGCACCTCCGCCCGCCCGCCGCGCGCGCGCGCGCGCGCAAG 350
OY 1984 caccatcggaacctcaagagttctctggaacctcaagggccggcggtgacctgag 2043
DB 349 GCGCGCCCGGAGACTGCCCGCCCGCCCGCCCGCCCGCCCGCGCGCGCGCGCG 290
OY 2044 tgggtgtactcgcgcgggggagactacatccgcgcgcggaagctgcgcagag 2100
DB 289 CAGTGTGCGCGCGCGGCTCTGCCAGCGCCGACCTGGAGCGGAAGTGAAGCGAAG 233

RESULT 12
T58686/c
ID T58686 standard; DNA; 5821 BP.
AC T58686;
DT 17-FEB-1998 (first entry)
DE DNA encoding S. friedlae tylosin biosynthesis gene products.
KW Tylosin; biosynthetic gene product; production; antibiotic; ds.
OS Streptomyces fridlae.
FH Key Location/Qualifiers
FH cds complement (73..1260)
FH /*tag= a
FH /product= tylE
FH complement (1277..1957)
FH /*tag= b
FH /product= tylD
FH complement (2286..4049)
FH /*tag= c
FH /product= tylH
FH 4094..4927
FH /*tag= d
FH /product= tylF
FH complement (5046..5660)
FH /*tag= e
FH /product= tylJ

US5672497-A.
PN 30-SEP-1997.
PD 30-DEC-1995; 575843.
PR 12-MAY-1989; US-351350.
PR 21-MAR-1986; US-842330.
PR 25-JUL-1986; US-890670.
PR 24-FEB-1987; US-018237.
PR 06-AUG-1991; US-742222.
PR 28-JUL-1993; US-107232.
PR 17-FEB-1994; US-198672.
PR 21-DEC-1995; US-575843.
PA (ELIL ) ILLI & CO. ELI.
PI Cox KL, Fishman SE, Hersberger CL, Seno ET;
DR WPI; 97-488860/45.
DR P-PSDB; W33272, W33273, W33274, W33275, W33276.
PT DNA encoding Streptomyces fridlae tylosin biosynthesis gene products
PT - for increasing tylosin production in Streptomyces spp.

```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 13, 2000, 02:55:32 : Search time 1880.46 seconds

(without alignments)
5727.858 Million cell updates/sec

Title: US-09-358-755-3

Perfect score: 2519
Sequence: 1 gcgcgcgcgcgcgcgcgcctc.....ctgatatacaataagcccaaaa 2519

Scoring table: IDENTITY_NUC

Searched: 780561 seqs, 2137953050 residues

Database: GenEmbl.*

Word size: 0

Number of hits that pass the threshold: 1561122

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl12:*
10: gb_pl2:*
11: gb_pl3:*
12: gb_pl3:*
13: gb_pl3:*
14: gb_sy:*
15: gb_un:*
16: gb_vl:*
17: em_fun:*
18: em_htg:*
19: em_hum1:*
20: em_hum2:*
21: em_hum3:*
22: em_om:*
23: em_or:*
24: em_ov:*
25: em_pat:*
26: em_ph:*
27: em_pl:*
28: em_ro:*
29: em_sts:*
30: em_sy:*
31: em_un:*
32: em_vl:*
33: gb_htg1:*
34: gb_htg2:*
35: gb_in1:*
36: gb_in2:*
37: em_ba1:*
38: em_ba2:*
39: em_hum4:*
40: em_hum4:*
41: gb_pr4:*
42: gb_htg3:*
43: gb_htg4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2378	94.4	2877	41	AF067864	AF067864 Homo sapi
2	713.4	28.3	227968	11	AF053356	AF053356 Homo sapi
3	550.4	21.8	580	11	HUMR50H04	AF085928 Homo sapi
4	303.8	12.1	2500	4	GGTFR	X55348 G.gallus mR
5	279.2	11.1	5010	9	HSTR	X01060 Human mRNA
6	278.6	11.1	2826	10	HUMTFR	M1507 Human trans
7	254.4	10.1	2460	12	CRUTRANREC	L19142 Cricetus
8	245.4	9.7	3413	12	RATFR	M58040 Rat transfe
9	213.4	8.5	2292	12	MTRRMRNA	X57349 M.musculus
10	158.6	6.3	780	5	I23809	I23809 Sequence 27
11	146	5.8	250	13	G14981	G14981 human STS S
12	125.4	5.0	540	5	I23811	I23811 Sequence 29
13	103.6	4.1	2570	12	AF009921	AF009921 Rattus no
14	91.6	3.6	2320	10	HSAN012371	AJ012371 Homo sapi
15	87.2	3.5	660	5	I23810	I23810 Sequence 28
16	73.4	2.9	3152	10	HSAN012370	AJ012370 Homo sapi
17	67.8	2.7	1428	12	AF039707	AF039707 Rattus no
18	67.8	2.7	2348	12	AF040256	AF040256 Rattus no
19	67.8	2.7	2899	12	RN075973	U75973 Rattus norv
20	66.8	2.7	2603	12	AF026380	AF026380 Mus muscu
21	57.8	2.3	1278	1	ABIPDAGEN	X88853 A.brasiliens
22	56.6	2.2	43632	1	SC6A5	AL049485 Streptomy
23	56.2	2.2	2532	3	AF050502	AF050502 Sus scrof
24	56	2.2	22449	1	SPSNBCDE	Y15548 S.pristinae
25	56	2.2	22449	1	SPSNBCGEN	X98690 S.pristinae
26	55.6	2.2	306	13	G45215	G45215 221292.1 ze
27	54.8	2.2	189	12	S81327	S81327 transferrin
28	54.6	2.2	3041	12	M0STRANS01	L13369 Mouse delta
29	53.8	2.1	2653	5	I23794	I23794 Sequence 1
30	53.8	2.1	2653	10	HUMPSM	M99487 Human prost
31	53.8	2.1	2518	41	AF176574	AF176574 Homo sapi
32	53.6	2.1	4459	1	TTPEST	X65609 T.thermophi
33	53.6	2.1	7621	1	TTPESTRO	Y15464 Thermus the
34	53.6	2.1	4171	1	TTPESTG	Z12118 T.thermophi
35	53	2.1	38995	1	SCF34	AL109974 Streptomy
36	51.8	2.1	3508	2	AF069761	AF069761 Halobacte
37	51.8	2.1	2014	2	AZSIDP	L26240 Azospirillum
38	51.8	2.1	982	7	ZMHSP18K1	X54075 Maize mRNA
39	50.2	2.0	10295	1	D63799	D63799 Thermus the
40	50.2	2.0	40745	1	SC151	AL109848 Streptomy
41	49.8	2.0	24700	1	SCB8	AL033654 Streptomy
42	48.6	1.9	40105	1	SC5F2A	AL049587 Streptomy
43	48.6	1.9	843	7	ZMHSP18K2	X54076 Maize mRNA
44	48.2	1.9	4933	1	ABIPDC	X99587 A.brasiliens
45	48	1.9	1800	16	HS1UL3	M19121 Herpes simp

ALIGNMENTS

RESULT	1	LOCUS	AF067864	DEFINITION	AF067864 2877 bp mRNA	PRI	27-JUL-1999
ACCESSION	AF067864	NID	95596369	VERSION	AF067864.1	GI:5596369	
KEYWORDS		SOURCE	human.	ORGANISM	Homo sapiens		
REFERENCE		1	(Dases 1 to 2877)				
AUTHORS			Kawabata,H., Yang,R., Hirama,T., Vuong,P.T., Kawano,S.,				
TITLE			Molecular cloning of transferrin receptor 2. A new member of the				
JOURNAL			J. Biol. Chem. 274 (30), 20826-20832 (1999)				

Dh	1992	GAACCTCAACGAGATTCTCTGGGACCTCAAGGCCCGGGGCTGACCTTCGCAATGGGTGTA	2051
QY	1694	ctcgcgccggggggaactacatccgggcgggcgnaaaagctcggcgcaagatctacagtc	1753
Dh	2052	CTCGGCGCGGGGGACTACATCTCCGGCGGGGAAAAGCTGCGGACGAGATCTACAGCTC	2111
QY	1754	ggaggaagaagagacgagcgaactgcacagcatgtcacaagctgagcaataagcggtgagtc	1813
Dh	2112	GGAGAGAAAGAGACGAGGAGCTGACAGCGATGTATCAACAGTCGCCATAAATGCGGGTGAAGTT	2171
QY	1814	ctacttccttcccaagtagtgcgcgaagcagctcccgcttcgagcaatcttaatgg	1873
Dh	2172	CTAATTCCTTTCCAGTACGTGTGCGCAGCCAGCTCCCGTTCCGCGCAATCTTATGGG	2231
QY	1874	ccgtgtagaacacacagctctgggcgcacctgtctgacacacctcgtgcgtctgcgtccacaag	1933
Dh	2232	CCGTGGAGACACACAGCGGTGGGGCCCGTGGTGAACACACTGGGGCTGTGCGTCCACAG	2291
QY	1934	ctccggagacccccgggggacacctctctccactggtcttcagagagccgtttccggcgta	1993
Dh	2292	CTCCGGAGACCCCCGGGGCCACCTTCCTCCATGCGCTTCCAGGAGAGCGCTTCCGGCGTCA	2351
QY	1994	gctagccctgtctaacctccggagcgtctgaaggggagacaaatgctgttagcggtgagtcgtg	2053
Dh	2352	GCTTAGCCCTGCTACCTTGGACGCTGCAAGGGGGAGCAATGCGCTTAAAGGGGAGTGTG	2411
QY	2054	gaacattgataacaactctctgagcccttgggagatctccacatcccgcttccccagtaag	2113
Dh	2412	GAACATTGATTAACAACACTTGTGAGGCCCTGGGAGTCTCTACATCTCCCGTCCCAATCAAG	2471
QY	2114	agctctctctgtctcctcgctgtgaatgataaggtcagggaggttggctcagagtaacactc	2173
Dh	2472	AGCTCCTCTCTCTCTCTCTGTAATGATTCAGGGGTCAAGGAGGTGCTCAGATCCACCTC	2531
QY	2174	tcatctgtgtgtaaatcttcatataccctcaacatctctccagagccccagaccgcgc	2233
Dh	2532	TCAATTCGTGATCAATTTCTATTATCCCTTACACATCTCTCCAGGAGGCCAACACCCAGC	2591
QY	2234	acagatactcacacacaccccaagccctcgtgagtgtagctgacccctaagtgtgacgtctactg	2293
Dh	2592	ACAATATTCACACACACCCAGCCCTCGAGTGTAGCTGACCCCTAATGTAGCGTCACTAGT	2651
QY	2294	tcggttaatcaagagtagtagcatccctctcaatcaacagccccctctcccttcttgggtctc	2353
Dh	2652	TGCGTTATATCAGAGAGTAGATCCCTTCAATCAACAGCCCTTCCCTTCTTGSGGTCTC	2711
QY	2354	catatccatagagacacactcttggggaggtttgcttaggcccttggagactcggcgcaagctcgttag	2413
Dh	2712	CATATCCTTAGAGACCACTCTGGAGGTTGTGTAGGCCCTTGGAACCTGGCCAGCTCTGTTAG	2771
QY	2414	tgtagaagagatcgctcgtgacacatagccttaatgagccacaggtgtctgtgtgtgaagggac	2473
Dh	2772	TGGGAGAGATTCGCTGCGACCATATGACCTTAATGGCCAAACAGTGTGTGTGTAAGGGGCG	2831
QY	2474	gtggagattcaatatcaataaacaccactgatalcaataaagccaanaa	2519
Dh	2832	GTGGAGTTTCATATCAATTAACCCACCTGATATCATATAGCCAAA	2877
RESULT	2		
AF053356			
LOCUS	AF053356	227968 bp	DNA
DEFINITION	Homo sapiens chromosome 7q22 sequence, complete sequence.		PRI
ACCESSION	AF053356		24-NOV-1998
NID	93135305		
VERSION	AF053356.1	GI:3135305	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 227968)		

AUTHORS	Glockner,G., Scherer,S., Schattevoy,R., Boright,A., Weber,J., Tsui,L.C. and Rosenthal,A.
TITLE	Large-scale sequencing of two regions in human chromosome 7q22: analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci reveals 17 genes
JOURNAL	Genome Res. 8 (10), 1060-1073 (1998)
MEDLINE	99018118
REFERENCE	2 (bases 1 to 227968) Gloeckner,G., Rosenthal,A., Drescher,B., Weber,J. and Schattevoy,R. Direct Submission Submitted (05-MAY-1998) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
AUTHORS	
JOURNAL	
MEDLINE	
REFERENCE	
JOURNAL	
FEATURES	Location/Qualifiers
SOURCE	1..227968 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7q22" complement(1..197) /note="SINE/Alu, pos. , in, rep.: , 2..197"
repeat_region	/rpt_family="Aluo" 1..169722
misc_feature	/note="PAC clone 3763" complement(join<309...578,2314...2458,3620...3831, 8786...9180,12095...12222,12821...12985,13079...13326, 14784...14928,15156...15357,16624...16804,18675...18769, 19928...20039,20325...20562,21131...21378,21722...21863, 22834...23036,24183...24361,24472...24582,26030...26173, 27852...27997,28543...28667,29738...29857,30045...30186, 32978...33158,35152...36676,37121...37205,37491...37762, 41053...41140,41322...41593,41773...41919,46462...46608, 46717...46795,47291...47382,48230...>48487)) /product="zonadhesin" complement(309...578) /note="100.000, excellent"
exon	complement(join<309...578,2314...2458,3620...3831, 8786...9180,12095...12222,12821...12985,13079...13326, 14784...14928,15156...15357,16624...16804,18675...18769, 19928...20039,20325...20562,21131...21378,21722...21863, 22834...23036,24183...24361,24472...24582,26030...26173, 27852...27997,28543...28667,29738...29857,30045...30186, 32978...33158,35152...36676,37121...37205,37491...37762, 41053...41140,41322...41593,41773...41919,46462...46608, 46717...46795,47291...47382,48230...48383)) /codon_start=1 /product="zonadhesin" /protein_id="AAC78790.1" /db_xref="PID:g3135306" /translation="MLDPPKRNARPGQAVLLSPVLSGGTSFSEHYILRGSGALL HIYASGVIRKHTLFESGOPPMNQVAIVYTATLRGFAPVGFGFKTEPRPAVDNLT SIACGGEFRCPDEPNADHPRCMVQISCGGHAKHGKNKGPCAGCAPAANVTIT TCCDFEDDAKPLCDMSGVSDADEDMWVASGPSPTSGTAGGCTPNGGCSLIHNESNF HRGGVALTLSPDLMEQGCLVFRAHHNEFTSWAQGLLDLALSIRBSCNGNHAIYELE QRSFWMLTVTPAGTFPLRLFMHEGRSTAAVLDDALSSIRBSCNGNHAIYELE ADESSHAGGSVRLSRFCAGDCIYCFEAFAMGTGTCMELLIGDPASPPIPKMK RVSGORPYOMNTSVTPSGHOQMOLFEKFAHMGICNTASYVMGILLNPGCPEVKLV ELPVPSVSSSTGSETGLTENPTISRKKPVSLERKSVTETKYREKRTIPTEKP TIESTKPTISEKNNMPSEKPTIPSEKPTILEKPTIPSEKPTIPSEKPTISTEKPV PTEPTPTEETTMYEPVIPTEKPSIPTEKPSIPTEKPSIMSELTISTEKPTICPP EKPTIPTEKPTIPTEKTISTEKPTTEKPTIPTEKPTIPTEKPTIPTEKPTISPERS LIAITEKLIIPTTEKPTIPTEKPTISTEEPTPRETTISTEKPSIPTEKPTIPTEET TSVETIISTEKTLIPMEKPTIPSEKPTIPSEKPTIPSEKPTIPTEKTLIPTKEPTIP IEERTISTEKTLIPTKEPTIPSEKPTIPSEKPTIPTEKPTIPTEETIISTEKLIPIRE KIPTSFKTLIPTKEPTIPSEKPTIPSEKPTIPSEKPTIPSEKPTIPTEKTLIARPH PSEPATGLAALVNSPAPSPTMTSVILGITTSNSSGMSCPPARKESCACSCKSS PRBCGPLCREGCVCNPFLESDNHCIOASSCNCYNNDYEPBAKFPTNCEJCHSCC WPBGRCOISOCIGTHITCOLKGOGYCYAGATCAITLYVGDIFYFDGNHFFKCVL CTYLIAOPGNSINDPFRTRANKNOEOGECVSLEYVAILPESTVLIAGRIIFVG QQVTLPIPSKVFLASGRFVELQTETGLRVRHDGDOQLYYVSSITYSKLGICLCN IDGNSDNDRHLKDGSPEGDKIELGNSQOTQDOEQOKRVQNVNSPSCDSLOSSMSG PGFCGRJLDHGPETICTLHVKAASFSDCMDBCGOGUOHLLCTHMSTMITTICODA

[illegible]

	misc_feature	Length=, 984, P=, 3.2e-14, S=, 70° /note="Region: Data base match" complement(475..567) /note="70, 1807:XYZ, 13441..13531, of, GBP L21998 HMMUC2X_1, Homo. sapiens, Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Homo, sapiens, intestinal, mucin, (MUC2), mRNA,, complete, cds... 1/98, length, =, 5179, P=, 1.5e-30, S=, 70° /note="Region: Data base match" 606..903
	repeat_region	/note="SINE/Alu, pos., in, rep.: 1..296" /rpt_family="AlusX"
	repeat_region	1004..1038 /note="83..30, %, 22, length: 2, #: 18, seq:, tt" /rpt_type="randem /rpt_unit=score: complement(1021..1313)
	repeat_region	/note="SINE/Alu, pos., in, rep.: 3..300" /rpt_family="Aluo" complement(1374..1684) /note="SINE/Alu, pos., in, rep.: 1..302" /rpt_family="Aluo" 1745..2050
	repeat_region	/note="SINE/Alu, pos., in, rep.: 1..302" /rpt_family="Aluo"
	Promoter	2026..2275 /note="7.13, TATA: -1, CAP: -1" 2166..2519
	Misc_feature	/note="14.60, 354, bp" /note="Region: GC Content" complement(2172..2421) /note="5.62, TATA: -1, CAP: -1" complement(2314..2458) /note="0.850, 0.45, 0.46, 0.56, 1.11, 0.49, 0.54, 0.53" complement(2314..2458) /note="94.000, excellent"
	exon	complement(2354..2464) /note="75, 26:xyz, 5749..5857, of, GBP U40024 SSU40024_1, Sus, scrofa, Eukaryotae;, mitochondrial, eukaryotes;, Metazoa; Sus, scrofa, zonadesin, (zan), mRNA,, complete, cds; The, amino, acid, sequence, predicts, seven, extracellular, domains,, including, a, mucin-like, domain, and, five, domains,.Homologous, to, the, D-do, .. ,'
	misc_feature	
	Query Match	28.3%; Score 713.4; DB 11; Length 227968; Best Local Similarity 99.9%; Pred. No. 2.1e-107;
	Matches 714: Conservative 0; Mismatches 1; Indels 0; Gaps 0.	
QY	1802	gcgggctggaagttctacttccttcccaagaagtgatcgccaggcagcattcccgtccgccaa 1861
Db	154517	GAGGTGGAGTCTCACTTCCTCCACAGATGTCGCACGCCAGCCTCCCCTCCGCCGA 154576
QY	1862	cattctcataggscgtlbgagaccacaagcttgggcgcctctgcctggacccaactctgcgctct 1921
Db	154577	CATTCTCATGGGCGCGTGAGACACAAGCTGGGCGCCCTGTGGACCACTCGGCTGCT 154636
QY	1922	gggtcccaacagctccggagaccgccgggggacacccctcccaatgcttccaagagagccg 1981
Db	154637	GGGCTCCAAACACTCCGGGACCCTCCGGGGCCACTCTCTCACTGGCTTCAGAGAAGCCG 154696
QY	1982	tttccggctcagctagacctgtctaacttgcagctgtgcaaggggcagccaatgaccttag 2041
Db	154697	T TTCGGGGGTACGATGACCCCTCTCACCTGACGCTGCAAGGGGCAAGCAATGGCCTTAG 154756
QY	2042	cggggatgtctggaacattgataaacactctgagggccctggggatcctcacatccccgt 2101
Db	154757	CGGGGATGTCTGGAACATTAATMAAATTTGAGAGGCCCTGGGGATTCCTCATATCCCCCT 154816
QY	2102	cccccaagtcaagaagctcctctgctctogcttgatagtatgaggttaaggaggggtgctc 2161
Db	154817	CCCCCAGTCAAAGATCCCTCTCTCTCGCTTGAAATGATTACGGGTTCAGGAGGAGGTGGCTC 154876
QY	2162	aagagtcacctctatctgtgtcataattctctaatcccttacacctacatctctccacggagc 2221

Db	Accession	Length	Insert	CDNA clone	YR50H04	29-AUG-1998
Db 154877	AGAGTCCACCTCTAATTCCTATCATATTTCTCATTTACCCCTACACATCTCTCCAGGAGC	154936				
Qy 2222	ccagagcccccagacacagatataccacacacccagccctgagtgatgacctgaatg	2281				
Db 154937	CCAGACCCCCACACACAGATATCCACACACCCAGCCCTGCAATGATGACCCCTATG	154996				
Qy 2282	acggtcacacgctgctgataatccagagatagatcctctcaatacaagccctccctt	2341				
Db 154997	ACGGGCTACATGTCGGTAAATCAGAGATGATCCCTTCATCATCACACCCCTCCCTT	155056				
Qy 2342	tctggggtcccccacatactagagaccactctggagaggttctgtagccctggacctgc	2401				
Db 155057	TCTGGGCTCTCCATACCTAGAGACACTCTGGAGGTTTCTGTAAGCCCTGGACCTG	155116				
Qy 2402	cagctctgtagtgaggagagatcgcctgacacatagctctatgaccaagtgctctgt	2461				
Db 155117	CAGCTCTGTATGATGGAGAGATCCGTGCGACACATGCTTATAGGCCAACAGTGTCTGT	155176				
Qy 2462	ggtgaagaaggcgctggagattccataataataaacaccctgatatcataagcca	2516				
Db 155177	GGTGAAGAAGGCGCTGGAGATTCCAAATCAATAAACCACCTGATATCAATAAGCCA	155231				
RESULT 3						
LOCUS	HUMYR50H04	560 bp	mRNA	PRT		29-AUG-1998
DEFINITION	Homo sapiens full length insert					
ACCESSION	AF085928					
NID	93483269					
VERSION	AF085928.1	GI:3483269				
KEYWORDS	FLI_CDNA.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia;					
AUTHORS	Eutheria: Primates: Carnivora: Canidae: Canis: Homo.					
	1 (bases 1 to 580)					
	Woessner, J., Tan, F., Maria, M., Kucaba, T., Yandell, M., Martin, J.,					
	Math, G., Bowles, L., Wylie, T., Bowers, Y., Steptoe, M., Theising, B.,					
	Geisler, S., Allen, M., Underwood, K., Chappell, J., Person, B.,					
	Gibbons, M., Harvey, N., Pape, D., Chamberlain, A., Morales, R.,					
	Schur, R., Rietter, E., Kohn, S., Swaller, T., Beymer, K., Hillier, L.,					
	Wilson, R. and Waterston, R.					
	Full Clone Sequencing of the Longest Available Member from Each					
	UniGene Cluster					
	Unpublished					
	2 (bases 1 to 580)					
	Waterston, R.					
	Direct Submission					
	Submitted (24-AUG-1998) Department of Genetics, Washington					
	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA					
	SUBMITTED BY:					
	Genome Sequencing Center					
	Department of Genetics					
	Washington University					
	St. Louis MO 63108, USA					
	http://genome.wustl.edu/gsc					
	mailto:estevatson.wustl.edu					
COMMENT	<p>NOTICE: This sequence represents the full insert of this cDNA. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived. We have tried to obtain double-stranded, or double chemistry sequence across the entire clone, but potentially, there are areas in the sequence where this level of coverage was not achieved. Nevertheless, we are confident of the accuracy of this sequence as all regions of low quality, as defined by PHRAP (P. Green, in preparation), were visually inspected and edited accordingly. The consensus quality values for this sequence have been submitted separately.</p>					
	<p>SIMILARITY INFORMATION:</p> <p>probable match to Homo sapiens protein AF053356 (PID:g3135312)</p>					

FEATURES	The location of this clone is unknown.
Location/Qualifiers	
source	1..580
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone_lib="Soares_fetal_liver_spleen_1NPLS"
	/clone IMAGE:208739
misc-feature	1..510
	/note="probable match to Homo sapiens protein AF053356 (PID:g3135312)"
misc-feature	43..549
	/note="Similar to Homo sapiens protein P02786 (PID:g136378)"
misc-feature	43..549
	/note="Similar to Rattus norvegicus protein A34549 (PID:g112425)"
misc-feature	43..549
	/note="Similar to Mus musculus protein S29548 (PID:g346962)"
misc-feature	43..549
	/note="Similar to Cricetulus griseus protein A48592 (PID:g539790)"
BASE COUNT	120 a 198 c 168 g 94 t
ORIGIN	
Query Match	21.8%; Score 550.4; DB 11; Length 580;
Best Local Similarity	98.9%; Pred. No. 1.7e-80;
Matches 554; Conservative	0; Mismatches 6; Indels 0; Gaps 0.
QY	186 gccgcctcgtactcagaagaattcgcgcgcgcgtctctccgcacagaagctggaccagctgtgg 245
Db	1 gccgcctcgtactcagaagaattcgcgcgcgcgtctctccgcacagaagctggaccagctgtgg 60
QY	246 accgacacgcactacgtgtgggctgcgaattccggatccggctcaccaccaacacctgcaac 305
Db	61 accgacacgcactacgtgtgggctgcgaattccggatccggctcaccaccaacacctgcaac 120
QY	306 tgggtcgaatgaagccgggaaaggtctggagagacgcctgcgccttggaagaccctgaactcacc 365
Db	121 tgggtcgaatgaagccgggaaaggtctggagagacgcctgcgccttggaagaccctgaactcacc 180
QY	366 tggccctacagcgcacatcggcacaagctcagcgaagagctgtgtgtacgcccactacagcggcgg 425
Db	181 tggccctacagcgcacatcggcacaagctcagcgaagagctgtgtgtacgcccactacagcggcgg 240
QY	426 ccgcaagaacctgcaggaacctgcgggcacgaagggcgttgatccagctgggcgcgcctgtcgtg 485
Db	241 ccgcaagaacctgcaggaacctgcgggcacgaagggcgttgatccagctgggcgcgcctgtcgtg 300
QY	486 gttgcgctgggggtgatatgaactctggcccaagaagtgacaatgctcaggaacttgggggt 545
Db	301 gttgcgctgggggtgatatgaactctggcccaagaagtgacaatgctcaggaacttgggggt 360
QY	546 caaagagtgcataataccacagagcagcggactctccacagaccacccaagccaagc 605
Db	361 caaagagtgcataataccacagagcagcggactctccacagaccacccaagccaagc 420
QY	606 ctgtccacgccaagcagcagtgatgtacatgtgcaccttggaacttggaagacccttaccaca 665
Db	421 ctgtccacgccaagcagcagtgatgtacatgtgcaccttggaacttggaagacccttaccaca 480
QY	666 cctggcttcctctccttaataaaccacagttccctccagttgcatatcaagccttccc 725
Db	481 cctggcttcctctccttaataaaccacagttccctccagttgcatatcaagccttccc 540
QY	726 agcatccagccaccccat 745
Db	541 agcatccagccaccccat 560

```

RESULT 4
LOCUS GGTFR 2500 bp mRNA VRT 30-SEP-1991
DEFINITION G.gallus mRNA for transferrin receptor.
ACCESSION X55348
NID 963806
VERSION X55348.1 GI:63806
KEYWORDS transferrin receptor.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 2500)
AUTHORS Chan,L.N.L.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-1990) Chan L.N.L., University of Texas Medical
Branch, Department of Human Biological Chemistry and Genetics,
Univ. Texas Medical Br., Galveston, TX 77550 U.S.A
REFERENCE 2 (bases 1 to 2500)
AUTHORS Gerhardt,E.M. and Chan,L.N.L.
TITLE Structure and Organization of the Chicken Transferrin Receptor and
JOURNAL Its cDNA Sequence
FEATURES
source
unpublished
Location/Qualifiers
1..2500
/organism="Gallus gallus"
/db_xref="taxon:9031"
/cell_type="erythroblast"
/clone_lib="gt11,chicken erythroblast"
/clone="CTRBV4"
79..2415
/standard_name="CTRDNA1"
/codon_start=1
/product="chicken transferrin receptor"
/protein_id="CAA39035.1"
/db_xref="PIR:G63807"
/db_xref="GI:63807"
/translat="MDHARALSNLESVPMSTYRSIRQTDGNSHVMKLSADDE
EGGDIERPHMVAQPOKRNKRLFLVIAALLLIGLISYRRIEIAACQ
DSGGCEIPTASVLDGEGVEEIQGPVIFWPELKMLSKLVNLRVR
GVDSFAGAEIDNMATYIHEERFNLGVWDEHYIKQVRSSTINQVSI
ILETDAIVAVSRSQSGKPVYVNGLKDFEIIQKVVASINGTIVIRAGITLAE
KYANAEKAAAGVIMVDSLMGITDITLIPGHAIGTQDPYTPGSPRHOFPVE
SSGPHIAVQTISSAAARLFSKMDSDTSEBKGKAIHCKTKTESIMKLDVN
SMKDRKILNIFGAIQGFEEPDRTVIVGAQRDSQPVAAAGTALLLELARI
KNEGYPRRSIIIFASWSAGDVAVGATEWLEGSAMLAHAKAFYIILDAVGA
ISASPLVIMLIGSIRKGVNPAVSESLNRLGPDVKAAPPVGLDNPFLAYSGI
PVLSFGFYNKDEEYRFLDRTKGTLEMLRKIDNLDALAAAEVAGGAAALRHDHLE
LDIGRASEELAAQEEFLPYIKFVRELGLTLMLEFRAGDFOFAVVALRDIANSDE
NRYIRALNDRMKRVYEDFLSPILSKVDYFRIIFFGKPHILRSLEVHLQILKTR
SYDLNLRQDLALATWTINGRPMWVVISGKLTMSRH"
BASE COUNT 650 a 545 c 707 g 598 t
ORIGIN
Query Match 12.1%; Score 303.8; DB 4; Length 2500;
Best Local Similarity 52.1%; Pred. No. 1.2e-40;
Matches 832; Conservative 0; Mismatches 717; Indels 48; Gaps 5;

```

```

Db 917 ATGCCAAGAGCGACGACGCTGAGTCTCATGACT----- 957
Qy 586 aggaaccacccaagccaagcctgtccaagccaagcagtgatgacatgtgcacttgg 645
Db 957 -GGATTCACTACTATGATGGAATATACACTTATCCATTGCGACATGCCACTTGG 1015
Qy 646 gaacttgaagacccctacacacttgccttcccttcaataaaccacgttccctcag 705
Db 1016 GAACCTGGAACCCCTTACACCCAGCGCTTCCCTGTTCAACACACACAGCTTCA 1075
Qy 706 ttgcacatcagcgccttccagcattccagccacccaatcagtgagcattgcctcc 765
Db 1076 TTGATCTTCAGACACTACCCACATTCGTTTCAGACATCTCTACCAATGACG 1135
Qy 766 gctcttgaaggaagctcaagcctctgtgtcccccgaatgtgcagggagcctctag 825
Db 1136 GGCTGTTACAGCAAAATGATGAGACACATGCTGTAAGGTTGGAAGTGCGATCAT 1195
Qy 826 gct-----cccttacccttgcctcccgccagcagcagctgcgtcagtgatc 879
Db 1196 CCGTAAGGTGACACAAAGCAGGAGCAGACATATGCTGAACCTAGTGCACAAAT 1255
Qy 880 acagaccctccaccccccacacatcttcgctgcacagcgcgcagcagcagcag 939
Db 1256 CCAATGAAGACAGAGAGATTTGACATCTTCGCTCTATCCAGGATTTGAAGAC 1315
Qy 940 atcaactcagctgtcatcctcgggcccagaggaatgtatgtggccccaagagcag 999
Db 1316 ATCGGTATGTTGATTGATGAGACAGACACTCTGGGGCCAGAGATGGCTTAAG 1375
Qy 1000 ctgttgagcagctatcctcctgtgagctgtgtgcagccttccctcctcagtg 1059
Db 1376 GCACGTGAACCTGTATATTTGTTGGAACCTGCGGTGATTCAGACATGGAAGAA 1435
Qy 1059 -ggctccggcccccagagcagcctcctcctcacaagcggagcagtggtgactt 1116
Db 1436 AGGCGTACAAACCGACAGCAGACATATCTTGTGCTGCTGAGATGACAGAGACT 1495
Qy 1117 gctgtggtccacagcagcagcagcagcagcagcagcagcagcagcagcagcag 1176
Db 1496 CTGTGGGTCTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1555
Qy 1177 tgaactgtgagcctcagcagcagcagcagcagcagcagcagcagcagcagcag 1236
Db 1556 CTTACATCAGCTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1615
Qy 1237 cctctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1296
Db 1616 CTTGCTGTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1675
Qy 1297 ggcagacctctatgaacagcagcagcagcagcagcagcagcagcagcagcag 1356
Db 1676 CAGAGAGCCTCTATACACACTTGCAGCCAGACGAGGCTAAAGCAGTTGTT----- 1726
Qy 1357 ggcctcaccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1416
Db 1726 --CCTTGTGCTGATATGCAAGCTTCCCTTCTGCGCTGCTGCTGCTGCTGCTGCT 1783
Qy 1417 tcgagtcctccttataagcagcagcagcagcagcagcagcagcagcagcagcagc 1476
Db 1784 TGTCTTTTGTGTTCTCATATTAAGATGAGAAATATGCTTCCGCGACACTAAGG 1843
Qy 1477 ctatagaacctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1536
Db 1844 CACTGAGAACCTGAGAGAAATATGATAAT-----CTGATGCTCTTCTGCGCTG 1897
Qy 1537 ccaagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1596
Db 1898 CAGAACTACTCTGACAGACAGCCTCTCAGGCTGACCCATATCATGAGCTCTTCT 1957
Qy 1597 tcggcgcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1656

```


Db	1998	TGCGAGAGTACAGTAAGAATTTACTGGCATACCAGAGAGAGTTTTCCTTAGCAATTAAAG	2017
QY	1657	aactcaagcccgccggcgtgaaccctgcagttggttacctcgccgcggggaggatacatcc	1716
Db	2018	AAGTCGGGAGAGTGCGGTGACCTTGGACTGGCTGTTTTTCCCGGTGGTACTTCCAGC	2077
QY	1717	ggcgcgcggaagaagtctgcgcgcagagatctacagcttcggagagagagacgagcactga	1776
Db	2078	GAGCTGTACTGTGCATGTGAGAAAGACATTGCCAAMAGTGAACGGGAGAAACAGGTCATCC	2137
QY	1777	cacgatgatcaagctgscacataatgcgggttagtttactctctctcccagtaactcgt	1836
Db	2138	GCAGGGCCCCGTGAATACACAGATGATGAAGGTGATGATTAAGCTTCTGTCCCGTACTCT	2197
QY	1837	cgccagccgactcccgcttcgcgcacatcttcataatgagccgttgagaaacacacgcgtg	1896
Db	2198	CACCAAAGAGTGTCCTTTTGCCCACTCTTCTTTGGCAAAAGCCCCCACACACCTTCGGGA	2257
QY	1897	ccctgctgaccacactgcgcgctgcgtcgtcgtccaacag	1933
Db	2258	GTCGTGTGAGACATCTGCAGCTGTGAACAACCAACAG	2294
RESULT	5		
HSTR	HSTR	5010 bp	PRI
LOCUS	Human mRNA for transferrin receptor.		11-APR-1995
DEFINITION	X01060		
ACCESSION	937432		
NID	X01060.1	GI:37432	
KEYWORDS	transferrin receptor.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;		
	Primates; Catarrhini; Hominoidea; Homo.		
REFERENCE	1 (bases 1 to 5010)		
AUTHORS	Schneider,C., Owen,M.J., Banville,D. and Williams,J.C.		
TITLE	Primary structure of human transferrin receptor deduced from the mRNA sequence		
JOURNAL	Nature 311 (1987), 675-678 (1984)		
MEDLINE	85012743		
COMMENT	Data kindly reviewed (19-FEB-1986) by C. Schneider.		
FEATURES	Location/Qualifiers		
Source	1..5010		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	264..2546		
	/note="put. transferrin receptor (aa 1-760)"		
	/codon_start=1		
	/protein_id="CAA25527.1"		
	/db_xref="PID:g37433"		
	/db_xref="GI:37433"		
	/db_xref="SWISS-PROT:P02786"		
	/translation="MMDARSAFSLRFGEFLSTRTSLAQVGDNSHVEMKLAVDE EENADNNTKANVTAPKRCSGSIICYGTNAVIVFPLIGVMIGLCKGVETPTEBERLA GETSPVRPEPEDPEPAARLYWDLKRLSLEKSDSTDFTSTIKLNENSIVPRAGSC KDENALCALVENOEFEEKSKYWMROHFVKIOVGSAONSIVYDNGRLYLVPNGPG VVAASKATVTKGLVHANFGTKRPDEDLPVNGSIYVRAGKITTPAEKVMAESLNKA ICGLIYMDQTKEPIYNALSTFEFHNLHGTDGPYPYIPGFSPNHOTPPERSGSGLNPIP OTISRAAEKELGNWEGCDPSMDKTDSICRNAYTSBSKNKVLTVASVLEKILINIFGV IKGEVEPHIVVGAQRDAWPMPGAKSSGVALLLKLQMFSNFKVASPLLYLIEKIIEH ASWSAGDGSGVATGLEDNSGYLSHLKAFYIINDKAVLGTSNFVSASPLLTYLIEKI TMQNVKHPTVOGLFYDNMSMASKYVELLDNAAPFPFLAYSGIPAVSFCDTDYPLY GTYMDIYKELIERIDLEKRVARAEEVAGOVVIKTLHVELINIDYERINSOLLSEFR LVNRADIKEMGISLOWXVARAGEFAVLTSLTDFGAEKTEDTRFMKLDNRVRYRE YHFIAPYSRKESPFRHYFWGSGSHITPLENLKLRKONGAEFTLEFRQLATWT TIIGCANALSGDVWDIDNEF"		
	457..527		
	/note="put. transmembrane segment (aa 63-88)"		
	1014..1022		
	/note="pot. extracellular glycosylation site (aa 251)"		
	1212..1220		
	/note="pot. extracellular glycosylation site (aa 317)"		
	misc-feature		
	misc-feature		
	misc-feature		

misc-feature	2442.	.2450	
BASE COUNT	1399 a	977 c	1155 g 1479 t
ORIGIN			
Query Match	11.1%	Score 279.2;	DB 9; Length 5010;
Best Local Similarity	54.0%	Pred. No. 1e-36;	
Matches 780; Conservative	0;	Mismatches 608;	Indels 56; Gaps 8;
QY	481	tcgtgtgtcgcgttgggggtgatcagctctgcgcgcgaagaagtgcacatgctcagagcttcg	540
DB	1027	tgattgtacagaccaggaataacacctttggcagaaaggattgcacaaatgctgaaacctttaa	1086
QY	541	gggtccaaaggagtgctcatataccagagccagcagcagctctccagagaccaccacaagc	600
DB	1087	atgcgaattgctgttggatattcatgagaccagacttaatt-----tc	1128
QY	601	caagcctgtccagccagcagcagcagtgatgtgcacatgtgcaccttggaaacttgagaccct	660
DB	1129	ccattgttaaacgacaaacttcatctttttggacatgctcatctggggagacagtgacctt	1188
QY	661	aacacacgtgctccctccctccatcaaaccccaattccctcagttgcatatcaagcc	720
DB	1189	acacacactggattcccttcccttcaatcacacacacttgcacatctcggtatcagagat	1248
QY	721	ttccccagcatcccaagcccaagcccatcagtgagacatgtccctcccgctgcgtggagagc	780
DB	1249	ttcccttaattatccttccagacaaatctccagagctgcctgcagaaaaagctgtttgggaata	1308
QY	781	tcaaaag-ccctgtgtgccccccaagaatgtgcagggagagcctccatgctccctatacac	839
DB	1309	tggaaaggagactgctccctctgactggaaaaacagactctacatgtagagatgtaacctcag	1368
QY	840	ctgggccccgggcccacgactgctgctagtgtgtgcaacaatacaaggaacctcccccacac	899
DB	1369	aaagcagagaat-----tgaaactcactgtgacccaatgtcctaamaagatataaaatt	1421
QY	900	aacacacatcttcgtgctgtcagaaaggccgctcagagccagatccactagctgtcatcg	959
DB	1422	cttaacatctttggagatttaataagcctttgtaagaaacagatcactatgtttactgtggg	1481
QY	960	gcccaaggagatgatcttgggccccagagagcagctaaactccgtgttggggagcgctatactc	1019
DB	1482	gcccaagagagatgcatggggccctcgagactgcaaaatccgggtgtaggcacacactctccta	1541
QY	1020	ctggagcgtgtgtcggaaactttccctcccatgtg---tgagcaaacgcttccggccccgaga	1076
DB	1542	ttgaaaccttgcccaaatgtttcagatataagcttttaaaagaatggggttttcagccacgcata	1601
QY	1077	agctccctctcatagcgtggagagcggtgtgtgacttttgagagcttgggtccacaggaatgtg	1136
DB	1602	acattattctttgccactgttgagtgctgagacttttgatccggttggtgctccactgaaatgg	1661
QY	1137	ctagaagcgtacctcagcgtgtgtcgaacctcaagccgctagtgtagcgttgacccaac	1196
DB	1662	ctaaaggagactcttgcgtccctgcatttaaaaggctttacattatattatgttgataaa	1721
QY	1197	gcagtgtcgtggggatgatacaagtttcaatgcacaaagaccagccccccttcgacaagtctatt	1256
DB	1722	ggcgattctttggtaacagacaaacttcaaaagttttgtccagccctactgtttatgcgttatt	1781
QY	1257	gagagtgctctgaagcagtgatgtctcccaaccacagatggggcgagaactctcatgaag	1316
DB	1782	gagaaacacatgcacaaatgtgaagcatcc---ggttactggccaatttttattatcagga-	1838
QY	1317	gtgtgtgttccaacatcccaagctggagatgtgtgagtatcgcgcctctaccatcgtgacgc	1376
DB	1838	-----cagcaactggggccacacaaagtgaactcactttagacamt	1880
QY	1377	atgtcctatctctcaagcgcttgttgggagtgccctgcgtctgagttctcccttatgag	1436
DB	1881	gcgtgtttccctttcttcacatattcttgaaatccacagcaggtttctgtttgtggcggag	1940

Db 1635 AAGCATCCGGT---TACTGGGCAATTCTATATACAGA-----CAGC 1673

Qy 1338 tggatgctgaagtgatccggccctaccatgacagcagtgctattctctcaagcc 1397

Db 1674 AACTGGCCAGCAAAAGTTGAAGAACTCACTTAACATAGCTGCTTCCCTTCTTGA 1733

Qy 1398 ttgtggagtcctgcgcgtcaggtctcctttagtagagacagcagccctaccatc 1457

Db 1734 TATCTTGGAATCCAGCAGCTTTCTTCTTTTGGCGAG---GACACAGATTATCTTAT 1790

Qy 1458 ctgcacacaaagagagacactatagaaacctgacataagtgctcgaagccgcctccc 1517

Db 1791 TTGGGTACACCATGACACCTATTAGGAAGTGAATGAGAGATTCCG---AGTTGAC 1847

Qy 1518 ggcgtggcccaagccgctggcccaagctcgcaagcagctcctcatccgctcaagcagat 1577

Db 1948 AAGTGGCAGCAGCAGCTGCAGAGAGCTGCTGATCAGTTCCGATTAACCTAACCATGAT 1907

Qy 1578 cgctgtgcccctcgaactcggcgtcgaagcgagcgctcgcctcgaagcacaatcggaac 1637

Db 1908 GTTAATTGAACCTGAGCATATGAGAGTACACAGCCACTGCTTTTCAATTTGAGGGAT 1967

Qy 1638 ctcaacagatctcttggagacctaagcccgcgagctgacccctgcagtgagggtactcg 1697

Db 1968 CTGAACCAATACAGACAGCAGATTAAGGAATGGGCTGAGTTTACAGTGGCTGATTTCT 2027

Qy 1698 ggcgaggggagactacatccggcgcggaagaaagctgcgcgagagatctacagctcgag 1757

Db 2028 GCTGTGGAGACTTCTTCCGCTACTCTCCAGACTAACACAGATTTCCGGGAAGCTGAG 2087

Qy 1758 gagagagagagagagacacgacatgacagctgacagatgacagatgagtgagttctac 1817

Db 2088 AAAACAGACAGATTTGATATGAAGAACTCATATGCTGATGAGAGTGAATATAC 2147

Qy 1818 ttccctccagtagtgcgtccagcagcagctcccgctccacacatcttaatggccgt 1877

Db 2148 TTCTCTCTCCCTACGATATCTCCAAAGAGTCTCTTCCACATGCTTCTGGGGCTCC 2207

Qy 1878 ggagacacacagctggcgccctgtgagacacactgcgctgc 1920

Db 2208 GGCTCTCACAGCTGCAGCTTTACTGTGAGAACTTGAACCTGC 2250

RESULT 7

CRUTRANREC 2460 bp mRNA ROD 09-DEC-1993

LOCUS Cricetulus griseus transferrin receptor mRNA, complete cds.

DEFINITION L19142

ACCESSION L19142

NID 9304528

VERSION L19142.1 GI:304528

KEYWORDS transferrin; transferrin receptor.

SOURCE Cricetulus griseus cDNA to mRNA.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.

REFERENCE 1 (bases 1 to 2460)

AUTHORS Collins,J.F., Lai,A., Domingo,D., Fitch,M., Hatton,S. and Trowbridge,T.S.

TITLE YrF is the conserved internalization signal of the transferrin receptor, and a second YrF signal at position 31-34 enhances endocytosis

JOURNAL J Biol. Chem. 268 (29), 21686-21692 (1993)

MEDLINE 94012749

FEATURES

source Location/Qualifiers

1..2460

/organism="Cricetulus griseus"

/db_xref="taxon:10029"

/cell_line="CHO-K1"

/tissue_type="Ovary"

/tissue_lib="Stratagene catalog #93/910"

1..2274

/codon_start=1

/function="endocytosis"

/product="transferrin receptor"

/protein_id="AA03576.1"

/db_xref="PID:9304529"

/db_xref="GI:304529"

/translation="MMDQARSAISNLFGEPELSTYRFSLAROVGDNSHVEKLAUDE
EENTDNMKASVRRHRLNGRLCGFIIVAVIFELIGFMIGYLGCRPOKDVRAE
TEGNSIELIOENIIPSSRLYAKMLDKLSEKIDAEFPDITKLSOTSRBPSGOME
NLAYIENQFPRFKLSKVRDEHYHYIQKSGSAQAQAVIINNGSDLYENRGSTVA
YSKATVSGKLHANGTARKDEEDLKYPVNGSLVIVRAKTRFAERVANAQSNALGV
LIYMDQKFPVVEALSLFHAHLGTDPTTPGFSPNHTQEPSSGLSPVQI
SRKAABKIQNNETNCPSPMNTDSLCKLESQGINVLNVNVLKETRILNIGYK
FEEDRIVYGAORDAMGGAAGKSSVGTGLLKLQAFSDMVSRGSEKSPRSIIFASM
SAGDFGAVATEMELEGYLSLKAFPTYNLIDKLVCTGFRVSAPLVLTLEKMO
DVRHPIDGRPLIRDSNMISKVEDLSIDNNAFPFLAISGIPAFNFCENEDRYEDTN
LDTRYELIKVQPLNKNVRAAEVAQGIIFKLTHDELWDYDMYNNKILSEVRELNO
FRADIKAMLSLQMLISARGDEFRAISRITDFHNAEKTRFVAREINNRIMKVEYHF
LSPVSPRESFPHIFWGSQSHLTITVALVENLKLKQNSSAFNETLFRNLALATWTIQ
GVANALSGDIDINEF"

3'UTR

BASE COUNT 744 a 487 c 539 g 690 t

ORIGIN

Query Match 10.1%; Score 254.4; DB 12; Length 2460;

Best Local Similarity 54.1%; Pred. No. 1.2e-32;

Matches 712; Conservative 0; Mismatches 566; Indels 38; Gaps 8;

Qy 622 cagtgtatgacatgtgcacctggaactgagagacccctacacactgctccctcct 681

Db 878 CACTCTTGGACATGCTCATCTAGACATGATGATTCATACACCTCGGGTTCCATCTT 937

Qy 682 tcaatcaaacccagttccctccagttgcatcatcagagctctccagcatccagccagc 741

Db 938 TCAATCATCTACATGATCTCCCATCTCAGTATCAGGATTCATGATACCTGTTGAGA 997

Qy 742 ccaatcagtgagagatagctccctccgctgctgaggaagct-caaagccctgtggccccc 800

Db 998 CGATCTCAAGAAAGCTGCGAGAAAGCTATTTCAAAACATGAGAAACAACATGCTCTCTA 1057

Qy 801 caagaatggcagggagagccctcctagctccctcctatcacctgggccccggccagagctg 860

Db 1058 GTTGGAACACAGATTTTGTAGTAGTAGTCA-----CAGGGTATTAAGTGT 1110

Qy 861 cgcgtagtggtcaacaatcacaagacctcacccccaacaacatctcgcgtgcatc 920

Db 1111 AACCTCTCTGTGAACAACGCTGTAAGAAACAAGAACTTAAACATCTTGGGATATT 1170

Qy 921 gaagcgccgtccagagccagatcaagtgtcatcgggcccaagagagatgcatggcc 980

Db 1171 AAAAGATTGAGGAACCAACCGGATCATTTGTAGTAGTCCCGCAGAGATGCTGGGGT 1230

Qy 981 ccaagacagactaaatccgctgtgagagcgctatactcctgagctggcgagacctt 1040

Db 1231 CCTGCTCTGCAAAATACATGATGTGGAACAGCTTCTTGTGAACATTTGCCAAGCATTC 1290

Qy 1041 tccctcagtg--gagcaagcgctccggcccgccgagaaagctcctctcatcagctg 1097

Db 1291 TCAGATATGTTTTCAGAGAGTGGGTTTAAACCCAGAGAGCATTTATCTTGGCAGCTGG 1350

Qy 1098 gacggtgtgacttggaaagcgtyggctccaagagagtggttagaagctcctcagctg 1157

Db 1351 AGTCCAGAGACTTCGAGAGCTGTGGTGCACACGAGTGGTAGAGGATACCTTCACTT 1410

Qy 1158 ctgcacctcaagccgtagtgcgtgagcctgagacacagcagtgctggggagatgacaag 1217

Db 1411 CTGATTTGAAAGCTTCACTATATATCTGTGATTAATTTGCTTGTGATACAGAAAC 1470

Qy 1218 ttcatgccaagaccagccctctcgtacaagttcatgtgaagtgctcgtgaagcagtg 1277

Db 1471 TTCAAGGTTTCTGCCACCTCTATTATATACACTTATTAGAAAGACAAATGACGAGACTA 1530

Qy 1278 gatctcccaaccagtgggcgagctcctcatgaagaagtggtgtcaccaatccagc 1337

Db 1531 AGACATCAATGTA---TGGGAACACCTCTCTATCGAGA-----CAGCAATGTGATC 1578
 QY 1338 tggagatgctgaggtgatcgccctaccacatggaagcagtgctattcttcacggcc 1397
 Db 1579 AGCAAAAGTGAAG-----ATCTTTCCCTTGTGAGAGATGCTCTTCCCTTTCTTGCA 1629
 QY 1398 ttgttggaagctccctgctgagctctctctcttaagagagacacagagctccatc 1457
 Db 1630 TATTCGGAATCCAGCAGATTTCTTCTGTTTGTGAGATG---AGGACTATCTTAT 1686
 QY 1458 ctgcacacaaaggagagacactatgaaacctgcataggtgtctgcaagcccgcc 1517
 Db 1687 TTGACACACTAATTTGATACCTATGAAACATTAATCAGAAAGTTC---CTCAGCTGAAAC 1743
 QY 1518 gcccgtgcccagcgctgtgcccagctcgcagagcagctctcatccgctcagcaagat 1577
 Db 1744 AAATGCTTCTGTCACACAGCAGAGATAGCTGTCAGTTCTAATTAATTAACCTTACCCATGAC 1803
 QY 1578 cgcctgctgcccctgctgagctcgcctgagggagcgtgctcgaagcagcagcggaac 1637
 Db 1804 ATTGAACTGAACCTGACATATGATATGATATACACAAAGATCTGCTATTTGTGAAGAG 1863
 QY 1638 ctcaacagagctctctgtggagacccaagcccgcggtgagccctgagtggtgactcg 1697
 Db 1864 CTGAACCAATTCAGAGCAGATATCAAGGCAATGGGTCTGAGTCTACAGTGGCTATTTCT 1923
 QY 1698 ggcgggggagctacatcgcggcgcggaagctgcggcagagatctacagctcgag 1757
 Db 1924 GCTCGGAGACACTTTTCCCTGCTACATCTGACCTGACGCTGATTTCCATTAATGCTGAA 1983
 QY 1758 gagaagagcagagcactgagacgcagtgacagtgagcgaatgagggaggtctctac 1817
 Db 1984 AAACAACAGATTTTCTGTCGAGGGAATCAATCAACCGATTAATGAAGGATATCTCAC 2043
 QY 1818 ttcccttccagtagctgtgcccagcagctcccgcttcccacacatctcatggcgct 1877
 Db 2044 TTCTCTGACCCCTACGATATTCACAGAGATCTCTCTTCCGACACATCTTCTGGGCTTC 2103
 QY 1878 ggaagacacagctgtggcgccctgtctgagacacactgcagctgtgctcacaag 1933
 Db 2104 GGCTCTCACACTCTCAGCGCTTATGAGAGAACTTAAGACTTCTGTCAGAAACAG 2159
 RESULT 8
 LOCUS RatRFR 3413 bp mRNA ROD 27-APR-1993
 DEFINITION Rat transferrin receptor mRNA, 3' end.
 ACCESSION M58040
 NID 9207463
 VERSION M58040.1 GI:207463
 KEYWORDS transferrin receptor.
 SOURCE Rat adult 120 day old testis, cDNA to mRNA.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 3413)
 AUTHORS Roberts,K.P. and Griswold,M.D.
 TITLE Characterization of rat transferrin receptor cDNA: The regulation
 of transferrin receptor mRNA in testes in Sertoli cells in culture
 JOURNAL Mol. Cell. Endocrinol. 14, 531-542 (1990)
 FEATURES
 source
 1..3413
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /dev_stage="120 day old adult"
 /tissue_type="testis"
 1..1870
 /gene="transferrin receptor"
 <1..1870
 /gene="transferrin receptor"
 /codon_start=2
 /product="transferrin receptor"
 /protein_id="AAA4273.1"

/db_xref="pid:g207464"
 /db_xref="gi:207464"
 /translation="IEFTDILIKOLSONTYTPREAGSOKENLAVYIENLFHDFRSKY
 WRDEHYKIQVNSVSONLYTINSNSINDPEAEGYAFKAEVGLKLVHAFGRK
 KDFELNYSVNSLVIYVAGRTITFAEKRAKSNALIVLIMDEPTPEVADLOFP
 GHAFHGTDPYTPGPPSPHNPFPPOSQSGPSIPVOTISAPAEKLEFKMEGNCPPS
 MNISSECTLESQNONKVLNYNVLKERIRIINITGVIGYEPRRYIVYGQRAMCP
 CVAASVETGLLKLIAOVSDMISKDSRPSRSTIFASWAGDVGAGVETMLSGYLS
 SLHKAFYIILDRVVLQTSNFKVASPLATYLMGKIMQVKKP1IDGXYLRNSNWS
 KIEELSLDNAEPFLAYSGLPVSFCEDEDEYDYLGRKLDYELILOKVPOLNOMR
 TAAEAVGQFIIKLHDIELIDYEMYNKLSFMKDLQFKADIKMDGLSPOMYSAR
 GDYFRATSLTFDFHNAEKTRFPMREINDIMVEYFLSPYSPRSPRHIFWGS
 GSHTLSAYVELRLROKNITAFNETLFRNOLALNTWTIGVANNLSGDIWNIDNEF"
 2846..3007
 /note="putative iron regulatory region"
 misc-feature
 3184..3301
 /note="putative iron regulatory region"
 BASE COUNT 971 a 668 c 741 g 1033 t
 ORIGIN
 Query Match 9.7%; Score 245.4; DB 12; Length 3413;
 Best Local Similarity 52.6%; Pred. No. 3.4e-31;
 Matches 766; Conservative 0; Mismatches 626; Indels 65; Gaps 8;
 Db 623 agtgtatggaatgtgacactgagagacccctacacacactggtctccctc 682
 Db 475 ATTCTTTGGACATGCTCATCTGGAACCTGGGATCCATATACCTCGCTTCTT 534
 QY 683 caataaacacagcttccctcagctgcatcagccctccacacatccagccagcc 742
 Db 535 CAACCACTCTCAGTTTCGCCATCTCAGTCACTGATGCTGATTTGCTTTATACCTGTCAGAC 594
 QY 743 catcagtgagaacatgtccctcccgctgtgagagagcctcaagg-ccctgtggccccc 801
 Db 595 GATCTCAAGAGCTCGCGAGAAACCTATCAAAACATGAGAGAACTGTCCTCTAG 654
 QY 802 aagaatgagagggagcctcctagcctccctatacactgagcccgagccagcagctgc 861
 Db 655 TTGGATATATAGATTCCTCATGTATGAGCTGGAACCTTCA-----CAGATCAAAATGTGA 707
 QY 862 ggcctagtggtgcaatcaccagcagcctccaccccaacaacacatctggctgacgc 921
 Db 708 AGCTCACTGTGAACATGACTGAAAGAAACAGAAATCTTAACATCTTGGCGTTATTA 767
 QY 922 aagggcgcctcagagccagatcactacgtgtgcatcggggcccagagggatgcatggccc 981
 Db 768 AAGGCTATGAGGACACAGACCGCTACATTTGTATGAGAGCCAGAGAGACGCTTGGGGCC 827
 QY 982 cagagagacctaatacgcgtgtgggagcggctatacctcctgagctgtgtcgagaccttt 1041
 Db 828 CTGGTGTTGCGAAGTCCAGTGTGGGACAGGCTCTTCTGTTGAACCTTGCCCAAGATTTCT 887
 QY 1042 cctccatagtg---gagcaagggctcccgcccgccagagagtgctctctcttcaacgttgg 1098
 Db 888 CAGATATGATTTTCAAAAGTGGATTTAGACCCAGCAGAGGATTAATCTTGGCCACCTGGA 947
 QY 1099 acgggtgttacttctggaagcgtgggctccagcagagtgctgaagagcactcagcgtgc 1158
 Db 948 CTGCAGAGACATATGAGAGCTGTGGTCCGACATGATGGCTGAGGGGATACCTTTATCTT 1007
 QY 1159 tgcacctcaagcgtagtgatgagcctgagacacagcagtgctgtgggagatgaagt 1218
 Db 1008 TGCATCTAAAGGCTTCTACTTACATTAATCTGATTAAGAGTCTCTGGTACTACCACT 1067
 QY 1219 ttcatagcgaagcagccctctcgacaagtctatgtagaggtgtctctggaagcagtg 1278
 Db 1068 TCAAGGTTTCTGCGAGCCCTTATTAATACCTTAATGGGGAATTAATGCGAGCATTA 1127
 QY 1279 attctcccaacacagctgtgagcactctatgacagcagtggtgttcccaacacagcgt 1338
 Db 1128 AGCATCCGATTGA---TGGAAATATCTATATCGAAA-----CAGTA 1166

[illegible]

ORGANISM	Unknown.	Unclassified.
REFERENCE	1 (bases 1 to 780)	
AUTHORS	Israeli,R.S, Heston,W.D.W. and Fair,W.R.	
TITLE	Prostate-specific membrane antigen	
JOURNAL	Patent: US 5538866-A 27 23-JUL-1996;	
FEATURES	Location/Qualifiers	
Source	1..780	
BASE COUNT	197 a 182 c 201 g 176 t 24 others	
ORIGIN	/organism="unknown"	
Query Match	6.3%; Score 158.6; DB 5; Length 780;	
Best Local Similarity	56.9%; Pred. No. 5.2e-17;	
Matches 390; Conservative	0; Mismatches 266; Indels 29; Gaps 5;	
QY	630 ggaatgagcaatgcaatggaactggagagaccctacacacactggtctccctccttaataa	689
Db	17 GGACATGGCCACCTTGGAACTGGAGACCCCTTACACCCAGGCTTCCCTTCCTGAACAC	76
QY	690 accca---gttccctcaggttgatcatcagaagccctccagcattccagccagccacatc	746
Db	77 ACCCANNNGTTCACACCGATTGAATCTTACAGACTACCCACATTCGTCTCAGACCTC	136
QY	747 agtgcagacatgcttccctccgctgctgtagaagctcaag--ggccctggtgcccccaag	804
Db	137 TCTACACTGGACAGAGCCAGGCTGTTCAGCAAAATGATGGAGACACATGCTGTGANN	196
QY	805 aatgcagagggaagcctccttagagctccctatcacctg-----ggc 845	
Db	197 NGTTGGAAGGCGCATCCANNNTTCCCTGTAAAGTNNACNNACAAAGCAGGAGANN	256
QY	846 cccggcgccagcactcggctagtggtgtcaacatcacagagcctccaccccatcaaac	905
Db	257 GCCAGANTTAATGGTGAACCTAGATGTGACAACTTCATGAGAGACAGGAGATTCCTG	316
QY	906 attcttggtgcatgcaaggcgcgtcgaagccagatacctactggtgcatcgggcccaag	965
Db	317 ATCTTGGTGGTATCCAGAGGATTTGAAGAACCTGATCGGTATGTTGATTGGAGCCAG	376
QY	966 aggaatgcatggtgccccagagagcaataatccgtgtggtggagcgtatatacctctgag	1025
Db	377 AGAGACTCTCTGGGGCCCGAGAGTGGCTAAAGCTGGCATGGAACCTGATTTGTGGAA	436
QY	1026 ctggtggtggaacttctctccatggtgagcaac---ggcttcgggccccgagaagcttc	1082
Db	437 CTGGCCCTGTGATCTCAGACATAGTAAACGAGGGCTCAAAACCGAGCGGAAGCTC	496
QY	1083 ctctcatcagccttggagaggtggtgactttggaagcgtgggctccagagagtgtagaa	1142
Db	497 ATCTTGTGCTAGCTGGAGAGTCAGAGACACTACGAGAGTGTGGGTGCTACTAGATGG	556
QY	1143 ggtactcctcagcgtgctgacactcaagcgttagtgcgtga--ggcttgagacaagcag	1200
Db	557 GGGTACTCTGGCAGTCTCATATGCAAAAGCTTCACTTACATCANNGCTTGGATGCT	616
QY	1201 tgcctggggagagacaagttcatgccaagaccagccctctctgacaagtcatattgaa	1260
Db	617 TCTCTGGGAGCAAGCCATGTCAGATTTCTGTCCAGCCCTTCTGTATATGCTGTGGGA	676
QY	1261 gtgtcctgaagcagtgattcttc 1285	
Db	677 GTATTATGAAGGGGTGAAGAACTCC 701	
RESULT 11		
LOCUS	G14981	250 bp DNA
DEFINITION	human STS SHGC-14854.	
ACCESSION	G14981	
VERSION	G14981.1	GI:1131744
NID	91131744	GI:1131744
VERSION	G14981.1	GI:1131744
DATE	04-JAN-1996	

Db 77 TACCTTATGCCAACAGGTGGCCTGTGGTGAAGGGCGCTGGATTCAATATCAATAAA 18

VERSION AF009921.1 GI:2653778

SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 2570)
AUTHORS Shneider,B.L., Thevananther,S., Moyer,M.S., Walters,H.C., Rinaldo,P., Devarajan,P., Sun,A.Q., Dawson,P.A. and Ananthanarayanan,M.
TITLE Cloning and characterization of a novel peptidase from rat and human ileum
JOURNAL J. Biol. Chem. (1997) In press
REFERENCE 2 (bases 1 to 2570)
AUTHORS Shneider,B.L., Thevananther,S., Moyer,M.S., Walters,H.C., Rinaldo,P., Devarajan,P., Sun,A.Q., Dawson,P.A. and Ananthanarayanan,M.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1997) Pediatrics, Yale University, 333 Cedar Street, New Haven, CT 06520, USA
FEATURES
source location/Qualifiers
1..2570
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/issue_type="ileum"
/note="ileal peptidase"
/codon_start=1
/product="1100"
/protein_id="AA87644.1"
/db_xref="PIR:P06379"
/db_xref="GI:2653779"
/translation="MHMAKILGVIGAAALGLGIIILGHAIKPAKEPLASSVSDSDLDIALDSMGOLDASRIKREINRELSKEPHVATSARDELAVYLLIGRKDSAGDITAKYETVLVSESTEDPNSVEVNGVIFHSFOPEKLTGEQAPNLYQAYAVPFGTPKGPLVANGSDEDFKLEAGINIKGIALTRKGVSGAKAINAAHGVAVGLVYDPGDIINDKSLPNEFFPMWGLPESGVSRYEYFGDLPYLPYAHVSFRLDPHNISGFPPIPTPIGIEDAKNLICNLINPDSVAGSGEYKLGPGEPENGNEPAGSEKVSYNRLERNSNVLGIIGAVPDRVYVIGNRDSWVGAVDPSSTAVILERSVLGLTKRKGTWRPRRSIIIFASMGAREELISGTEPTEFLSKIDRYTYTIVDISVFNATIRKAGTPVOSVITRSATRELSAPSSGSLIDNIRITNRSFVYGAVPSMOTLQASDYASFIHFLGTTSMDLATYVDRSTSAKIPETHTADTEDVYKFLDPGFSHQAIVARTGASVLRSLDSFLPLNVDSEYTLQSFQAQENIGALBESHNLSLGPLTVAEKKAAALALNOHILFLQKSSPDLGVARNVNDMLLEAFILPRAFPEERYVSHVMAPRTASVATFPGLANVYARAQEIINSAEAMAEVERQLSLAVVALECAATLQPVYDL"
BASE COUNT 575 a 747 c 690 g 558 t
ORIGIN

Query Match 4.1%; Score 103.6; DB 12; Length 2570;
Best Local Similarity 50.8%; Pred. No. 3,4e-08;
Matches 402; Conservative 0; Mismatches 354; Indels 36; Gaps 5;

QY 902 caacatcttcggtcgtcgaagagccgctcgaagccagatcactcgtgtcgcgagc 961
DB 1083 CAACGTCCTGGGCAATATTCAGGGGCAAGTGCATGATGATGTATACGGAA 1142
QY 962 ccaagaggatgcatgggcccccaagagcagctaacccgctgtgtggaagcgtactactcct 1021
DB 1143 CCACCCGGGACAGCTGGGTACATGGGCGACAGTGCACCCAGCAGAGGCACTGCTCTCTT 1202
QY 1022 gggagcgggagcgagcctttccctcactgtgtgagcaagcgttc---cgagcccgaggaag 1078
DB 1203 GGGAGATCTCTCGAGTCTTGCGGACCTTGCTAAAGAAAGGCACTGGGGGCCCGGTAGATC 1262
QY 1079 tctcctcttcacagcttgagcagcgtgtgtgactttggaagcgtgtgctccagcagagtcgct 1138
DB 1263 CACCATATTTTGGCAGCTGGGAGCAGAGAGTTCGGGCTCATGCGCTCCACACAATTTCAC 1322
QY 1139 agaaagctacctcagcgtgtcgcactcaagccgctagtgactgtgagccttgagccttgagcgc 1198
DB 1323 AGAGGAGTTCCTCAGCAGCTGAGAGCGCACAGTGCATCATCAACGCTGGACATCTC 1382

QY 1199 agtgcctggggagatgacaagtttcatcagccaaagacagccccccttcgtgacaagtcctatiga 1258
DB 1383 TGTGTTCTCCAAATGCTACCTCCGGGCACAGGGGACACCCCTGTACAAAGGTATCTT 1442
QY 1259 gaggctcctgaagcagctgtgattctcccaaccacagctggcgagactcctatgaacaagt 1318
DB 1443 CCTGGCACCAAGAGATAGTGCACACGACGCTCCAGTGGCCCTCAGATATGACAAACG 1502
QY 1319 ggtgttccaacatcccaagctgtgagctgagtgatgcgcgcctaccacatgaaca--- 1375
DB 1503 GATCCGATATACCTAACCGGACAGTAGTCCAGTATATGAGACTGTTCCAGCATGGGCACT 1562
QY 1375 -----gcagtcctat---tccctcagccctttgtggagagcccgctgcagtt 1423
DB 1563 GGGTCAGGACGACGACTATGCTCATTTCACTTCTGGCAATCACCCTCATGAGACCT 1622
QY 1424 ctcccttatgagagc-----gaccagcctaccatcctcgtcacaca 1468
DB 1623 TGCGTACACCTATAGACCGGAGCAGACGACGACGATCTACCACTACACACAGC 1682
QY 1469 ggaagacacttatgagaacctgcataagtgctgcaaggcgccgtgccgttgccca 1528
DB 1683 GTTTGACAGCTTGACTATGTGAGAGATTTCTGGACCTGGTTTCAGCAGC---CACCA 1739
QY 1529 ggcgcctggcccaagcctcgcaaggcagcctcctcctcagcctcaacagcagatcgcctgtgc 1588
DB 1740 GGGTGTGGCCCGACTGTGGAGTGTGCTGCTCCGGCTGAGTGCACGCTTTCTCTGCG 1799
QY 1589 cctgcagctcgcgcgtcgaaggagcgtcctcctcagcagcagcagcagcagcagcagcagc 1648
DB 1800 CCTCATATGCTAGTACTATAGTACAGACTACAGAGTTTCTTCAACAGCGCTCCTCAAGAGA 1859
QY 1649 cctctggagacct 1660
DB 1860 CCTTGCGGCCCT 1871

RESULT 14
HSA012371
LOCUS HSA012371 2320 bp mRNA PRI 25-MAR-1999
DEFINITION Homo sapiens mRNA for NAALDase L protein.
ACCESSION AJ012371
NID 94539526
VERSION AJ012371.1 GI:4539526
KEYWORDS naaladase L gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2320)
AUTHORS Pangalos,M.N., Neef,J.M., Somers,M., Verhasselt,P., Bekkers,M., van der Helm,L., Fridpols,E., Ashton,D. and Gordon,R.D.
TITLE Isolation and expression of novel human glutamate carboxypeptidases with N-acetylated alpha-linked acidic dipeptidase and dipeptidyl peptidase IV activity
JOURNAL J Biol Chem. 274 (13), 8470-8483 (1999)
MEDLINE 99185063
REFERENCE 2 (bases 1 to 2320)
AUTHORS Pangalos,M.N.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) Pangalos M.N., Biotechnology, Janssen Pharmaceutica, 30 Turnhoutseweg, 2340 Beerse, BELGIUM
FEATURES
source location/Qualifiers
1..2320
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q12"
/tissue_type="small intestine"
/gene="naaladase L"
/gene="naaladase L"
/gene="naaladase L"
CDS


```

/function="peptidase"
/codon_start=1
/evidence=experimental
/product="NAALadase L protein"
/protein_id="CAB3968.1"
/db_xref="PID:e420518"
/db_xref="PID:94539527"
/db_xref="GI:4539527"
/translation="MONTKYLGLGALNALILGILGHFAIRKANSAPLDLEI
LETWQQLDNRILREMLRELSREPLHASSRDELDVQLLRNKPDEPSIDSEATY
EVLSPSEQPNVNDIVPTGGIHSCHRETBENYVDEQGPVQVQPYAAYASGTPQ
GLIVANRGAEDFEKELQGIKLEGIALTIRYGRGAKANAAHAGVAVLVTD
PADINDLSSPDETPPNWYLPSPGVRSYVYFPGDLPVLPAYSPSRVLAVS
GEPRIPTOPRGODARDLCLNGLTAPATMOALCHYRLGCFRDPGFPADSOVN
VSYRNLELRNSSNVLCITRGAVPDRYVLYGNHRDSWHAQVDPSSGTVLLELSV
LGLTLKGTWRPRRSIVFASWGAEEFGLISTEFTBEFKLQERTVAYINVDISFA
NATLRVQGTFFPVSIVFASWGAEEFGLISTEFTBEFKLQERTVAYINVDISFA
AGSDYAFVHFLGISMDIAYIDRSKTSARIPTYTAEDTIDYDKFLDPFSSHO
AVARTGSVILRLSDSFELPLKYSSETLRSFLQAAQODLGLAEHDSISLPIVTA
VEKEEAALIGORISLTOKGSPDPIQVRLNDOQMLEFTIPNPRAPFERYSHTV
MAPSHGIRSHITPELSNACSRARDTASGEAMAEVQRLSTIVTALBGAATLRPADL
"
BASE COUNT      453 a      725 c      671 g      470 t      1 others
ORIGIN
Query Match      3.68; Score 91.6; DB 10; Length 2320;
Best Local Similarity 46.68; Pred. No. 3.1e-06;
Matches 672; Conservative 0; Mismatches 674; Indels 96; Gaps 8;
QY 314 tgaagccgggaaggtcgagagagcagctgcccgtggagagaccctgactctctcccta 373
D 415 TGAGGAGACCTGACCCGGGAGCAAGGGGCGCAGATGCTTACACCTATVGTCTCTTA 474
QY 374 caagccatcggcaagctcaagggagagctgtgtacgcccactacgggcccgaaga 433
D 475 TGCTCTTTGGAAACCCACAGGGCCTCTGCTATGCCAACCAGGGGCGGGAAGA 534
QY 434 cctgcagagacctgcgggcccagggcgttgatccagctgggcgcctgtctgtgcgct 493
D 535 CTTTAAAGGAGCTACAGACTCAGGGCATTCAAACTTGAAGGCCATTCGCTGACTCGATA 594
QY 494 gggggatgaagctgcggcccaaggtgacaaatgctcagaacttcgggggtcaagagt 553
D 595 TGGGGGTGTAGGGGCTGGGCCAAGGCTGTAGACGCTGCCAAGCAGGGGAGCTGGGCT 654
QY 654 gctcatataccagagcagcagcgtctccagagaccaccacaagcacaagcctgtccag 613
D 655 GGTGTGTACACAGACCTGCCGACATCAAGATGGGCTGAGCTACACCCGAGAACTT 714
QY 614 ccagcaggcagtgta-----tggaatgtgcaactggaactggagac 656
D 715 TCCCACTCTCGTACTACGCCCCCTCAGAGTGAAGGAGGAGGCTCTACTACGATATT 774
QY 775 TGGGGACCTTGACTCCCTACCTTCCAGCCGCTCCCTCTTCCGCGTGGACCTTGC 834
QY 835 CAATGTCTCCGCGATTTCGCCCAATTCTTACACAGCCCTTGGCTTCCAGATGCAAGA 894
QY 895 CCTGCTCTGTAACTCAACGGAAGACTTGGCCAGCCACCTGGCAGG---GAGCACTGG 951
QY 951 ctcccatatcactgggccccgggcccagac----- 859
D 952 CTGCCACTACAGGTGGGTCCGGCTTCGAGCGAGACTTCCAGAGACAGCA 1011
QY 1011 --tcgcgctagtggtcaacaatcagagagctccacccccaatcaacaatcttcggtg 916
D 1012 GGTGAATGTAGCGCTTACACCGCGCTGAGACTGAGAACTCTTCCAGAGCTCTGGCAT 1071

```

```

QY 917 catcgaagccgcctcaagccagatacctacgttgtatcgtgggcccagagagatgcatg 975
D 1072 CATCCGTGGGGCTGTGTGAGGCTGATCCCTGATGCTTATGGGAACCCAGGACACTG 1131
QY 977 gggcccaagagcagcctaaccgcgtgtgaggagagcctactccttgagctgtgaggac 1036
D 1132 GTTCCACGGGGCTGTGAGACCCACAGACTGGCACCAGCGCTCTCTGTGAGCTCTCCG 1191
QY 1037 ctttcctccatctgtgtgacaaagg---cttcggcccccgagaagctcctctcatcag 1093
D 1192 CCTGGGAGACCTGTGTGAAGAAGGACACTGGCGTCTCCGAGATCAATCGTGTGGAG 1251
QY 1094 ctggagcgtgtgtgaccttggaaagcgtgtggtctcagagatgtgctagaagctacc 1153
D 1252 CTGGGGGGCTGAGAGATTGGGCTCATTTGCTCCACGGAATTCAACAAGAGTCTTCAA 1311
QY 1154 cgtgtcacctcaagccgtagtgatcgtgagcctgagacaagcagtgctgaggagatga 1213
D 1312 CAAGCTCAGAGAGCGCAGCGTGGCTTACATCAACGTGACATCTCGGTGTTGCCAACGC 1371
QY 1214 caagttcatgccaagaccagcccccctctgacaagctcatgtgagatgtccctgaaga 1273
D 1372 TACCCTTAGGGGTGAGGGAGCGCCCTGTGCAAGGGTGTCTTCTGCAACCAAGA 1431
QY 1274 ggtggtatctcccaaccacagctggcagactctctatgaacagtggtgttaccaatcc 1333
D 1432 GATCCGCTTACCAAGCCCTGCGACCTGAGACTGACAGCAACTGATCCGGTACTTCAA 1491
QY 1334 cagcttgagtgatgtgagtgatcggccccctaccatgtga-----cagcag 1378
D 1492 CCGCAGACGCCGGGTGACGGGCTGTGCCCGAGTGGGTTCTGTGGTGCTGCGACGA 1551
QY 1379 tgcctatctctcagcgccttctgtgagagctccctgcagagctcctcttaagagga 1438
D 1552 CTATGACACCTGTCTTACCTTCTGGGACATCTCTCATGAGACATTCCTTATCTATGA 1611
QY 1439 c-----gaccagcctaccatctcctcgtcaacaagaagagactatga 1483
D 1612 CCGGAGCAAGACTTACGCCAGGATCTACCCACCTTACACACACCTTTGACACCTTGA 1671
QY 1484 gaacttcataagtgctgcaagcgcctgcgcgcgcgtgagcagcgtgtggccagcgt 1543
D 1672 CTATGTGACAAAGTTTGTGACCCGGGCTT---CAGACCCATCAGCTGTGGCCCGAC 1728
QY 1544 cgcagggcagctcctcctcagctcagcagcagatcagcgtcgtgcctcagacttcggcgcg 1603
D 1729 AGCGGGAGGTGTATTTCTCGGCTCAGTGAAGCTTCTTCTGCCCTCAAGTCAATGA 1788
QY 1604 ctacgggagcgtgctcctcagcagcagcagcagcagcagcagcagcagcagcagcag 1663
D 1789 CTACAGTGAAGACACTCCGACGCTTCTGTGAGGACGCCAGCAGAAATCTTGGGCCCTG 1848
QY 1664 gg 1665
D 1849 GG 1850

```

```

RESULT 15
LOCUS 123810 660 bp DNA PAT 07-OCT-1996
DEFINITION Sequence 28 from patent US 5338866.
ACCESSION 123810
NID 91603680
VERSION 123810.1 GI:1603680
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 660)
AUTHORS Israeli, R.S., Heston, W.D.W. and Fair, W.R.
TITLE Patent-specific membrane antigen
JOURNAL Patent: US 5338866-A 28 23 -JUL-1996.

```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 13, 2000, 19:03:57 : Search time 120.47 Seconds
(without alignments)
5231.459 Million cell updates/sec

Title: US-09-358-755-3
Perfect score: 2519
Sequence: 1 gcgtccgcgggagcgctct.....ctgatataaataagccaaaa 2519

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Word size : 0

Number of hits that pass the threshold : 623170

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	146.6	5.8	779	1	065535	Sequence used in 1
2	125.4	5.0	540	1	065537	Human gene signal
3	89.4	3.5	93	1	T19421	Sequence used in 1
4	87.2	3.5	660	1	065536	Sequence used in 1
5	53.8	2.1	2553	1	065520	Prostate-specific
6	53.8	2.1	2654	1	T36785	Prostate-specific
7	47.6	1.9	3059	1	N70566	Sequence encoding
8	46.2	1.8	1030	1	V99230	DNA encoding an ac
9	46	1.8	2353	1	037948	Sequence of a DNA
10	46	1.8	2353	1	V07530	Human transcriptio
11	45.8	1.8	53789	1	V21187	Amycolatopsis medi
12	45	1.8	4356	1	Q37543	Cardiac adenylyl c
13	45	1.8	4356	1	Q95540	Cardiac adenylyl c
14	44.8	1.8	6781	1	V45824	Maize phospholip
15	44.8	1.8	2135	1	X26879	DNA encoding a pro
16	44	1.7	1389	1	T18534	Chimeric herregulin
17	44	1.7	3576	1	V21186	Amycolatopsis medi
18	44	1.7	53789	1	V21187	Amycolatopsis medi
19	43.8	1.7	2846	1	V41260	Chlamydomonas rein
20	43.2	1.7	44377	1	T78508	Platenolide synth
21	43.2	1.7	44377	1	T80414	Platenolide synth
22	43	1.7	1278	1	Q21629	Bleomycin acetyl t
23	42.8	1.7	15079	1	Q91880	S. clavuligerus cl
24	42.8	1.7	8065	1	V38335	Manic-depressive i
25	42.8	1.7	6093	1	X25775	S. erythraea oleand
26	42.6	1.7	1542	1	T61551	Human Sox1 encodin
27	42.2	1.7	12001	1	076213	HSV L/ST region. H
28	42.2	1.7	4257	1	V10362	Infected cell prot
29	42.2	1.7	4257	1	V68520	The nucleotide seq
30	42	1.7	2729	1	N92629	DNA fragment conta
31	42	1.7	1164	1	N92631	ACY B2 gene - enco
32	42	1.7	1630	1	O66684	DNA fragment conta
33	41.8	1.7	15872	1	T68715	Streptomyces venez
34	41.6	1.7	2072	1	Q04529	Elastase gene. pse
35	41.6	1.7	2483	1	T12462	Human K+ channel 2
36	41.6	1.7	2483	1	V04874	cDNA sequence of t
37	41.6	1.7	1831	1	X26880	DNA encoding a pro
38	41	1.6	15464	1	O11415	Ryanodine receptor
39	41	1.6	1960	1	T43380	Human cytokine res

ALIGNMENTS

40	41	1.6	43280	1	T80413	Tylosone synthase
c 41	41	1.6	17955	1	V56642	Actinoplanes sp. a
42	41	1.6	7559	1	V82011	Maize phosphoenolp
43	40.8	1.6	2514	1	T30308	Mouse Sox-9 cDNA.
44	40.8	1.6	3934	1	T30309	Human Sox-9 cDNA.
45	40.6	1.6	5392	1	O64201	Sequence compislin

RESULT 1
ID 065535
AC 065535 standard; cDNA. 779 BP.
AD 11-JAN-1995 (first entry)
DE Sequence used in identification of PSM Ag cDNA.
KW Prostate-specific membrane antigen; PSM: prostate cancer;
KW transmembrane glycoprotein; imaging; targeting; tumour detection;
KW antibody detection; ds.
OS Homo sapiens.
PN WO9409820-A.
PD 11-MAY-1994.
PE 05-NOV-1993; U10624.
PR 05-NOV-1992; US-973337.
PA (SLOK) SLOAN KETTERING INST CANCER.
PI Fair WR, Heston WD, Israeli RS;
DR WPI: 94-167129/20.
PT Prostate-specific membrane antigen and DNA encoding it - is
PT useful for detecting haematogenous micro-metastatic tumour cells
PT and for identifying ligands which bind to PSM Ag
PS Disclosure: Page 119-120; 196pp; English
CC This sequence is given in the specification but is not mentioned in
CC the disclosure. The PSM coding sequence is useful for suppressing or
CC modulating the metastatic ability of prostate tumour cells to grow,
CC or for eliminating them. The protein is useful to identify or purify
CC ligands of the Ag. It is also an attractive target for Ab-directed
CC imaging and targeting of prostatic tumour deposits.
SQ Sequence 779 BP; 197 A; 182 C; 201 G; 175 T;

Query Match 5.8%; Score 146.6; DB 1; Length 779;
Best Local Similarity 56.8%; Pred. No. 1.2e-21;
Matches 389; Conservative 0; Mismatches 266; Indels 30; Gaps 6;

QY	630	ggacatgtcgaacttgggaacccctacacacacccctgctccctcctcaatcaa	689
DB	17	GGACATGCCCACTTGGAACTGGAGACCTTACACCCAGGC-TCCCTTCTTCAACAC	75
QY	690	accga---gtccctcagttgcatcatcaagcctcccaagcatcccaagcccatc	746
DB	76	ACCCANNNGTTCCACAGTTCATCTTACAGACTACCCACATCTCTTTCAGACATC	135
QY	747	agtcgacgaattgtccctccgctgctggaagctcaag--gccttgggcccccaag	804
DB	136	TCTAGACGTGCACACCGCAGCTGTTCAGCAAAATGATGAGACACATGCTTGAAAGN	195
QY	805	aatggcaagggaagcctcccaagcgtccctcattcaactg-----gac	845
DB	196	NGTTGGAAGGTGTCGATTCACANNNTTCTGTAGCTNNACNNACAAAGCAGAGANNNN	255
QY	846	cccgagccacgactcgcgtagtgtaacaacatcacaggaacctccaccccccaaac	905
DB	256	GCCAGANTATGCTGTAACATGATGTAACATTCATGAAAGCAGAAAGATTCTGAC	315
QY	906	attctggctgtaaggaagcgcctcaagagccagatcactagttctcggggccag	965
DB	316	ATCTTGGTGTCTATCCAGGGAATTTGAAGAACCTGATCGGTATTTGTGATGAGCCAG	375
QY	966	aggaatgatggcccccagagcactaaatccgcctgtggagcggctacacctggag	1025
DB	376	AGAGATCTCCTGGGGCCAGAGAGTGGCTAAAGCTGGCAGTGAACCTCTATTTGTTGGA	435

```

QY 1026 ctggtgcgacattctccatggtgagcaac---ggcttcggcccccgaagaatctc 1082
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 CTGGCCCTGTGATCTCAGACATAGTGAAGAAAGAGGCTACAAACCGAGGGAAGCATC 495
QY 1083 ctcttcacagctggagcaggtggtgacttggaaagcgtgggctccaggaatggtataga 1142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 ATCTTGCTAGCTGAGTGCAGAGAGACTACGAGAGCTGTGGGTGCTACTGAAATGCTGGAG 555
QY 1143 ggctaccacagcgtgtctcactcactcaaacgcgtagtgatgta--gcctggacaacgag 1200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 GGGTACTCTGCACCTGCATGCCAAACCTTTCACCTACATCANNCTGTGAGTCTCCAG 615
QY 1201 tgcctggagatgacagaattcattgcacaagaccgcccctctgacaagctccattgaga 1260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 616 TCCTGGAGCAAGCATTCATGAAGATTTCGCCAGGCCCTTGCTGTATATGCTGCTGGGA 675
QY 1261 gtgtctcgaagcaggtgattctcc 1285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 676 GTATTATGAGGGGGGTGAAGATCC 700

```

```

RESULT 2
Q65337
ID Q65337 standard; cDNA; 540 BP.
AC Q65337;
DT 11-JAN-1995 (first entry)
DE Sequence used in identification of PSM Ag cDNA.
KW Prostate-specific membrane antigen; PSM; prostate cancer;
KW transmembrane glycoprotein; imaging; targeting; tumour detection;
KW antibody detection; ds.
OS Homo sapiens.
PN WO9409820-A.
PD 11-NOV-1994.
PR 05-NOV-1993; U10624.
PR 05-NOV-1992; US-973337.
PA (SLOK ) SLOAN KETTERING INST CANCER.
PI Fair WR, Heston MDW, Israel RS.
DR MPI; 94-167129/20.
PT Prostate-specific membrane antigen and DNA encoding it - is
PT useful for detecting haematogenous micro-metastatic tumour cells
PT and for identifying ligands which bind to PSM Ag
PS Disclosure: Page 121; 196pp; English
CC This sequence is given in the specification but is not mentioned in
CC the disclosure. The PSM coding sequence is useful for suppressing or
CC modulating the metastatic ability of prostate tumour cells to grow
CC or for eliminating them. The protein is useful to identify or purify
CC ligands of the Ag. It is also an attractive target for Ab-directed
CC imaging and targeting of prostatic tumour deposits.
SQ Sequence 540 BP; 152 A; 106 C; 133 G; 147 T;

```

```

Query Match 5.0%; Score 125.4; DB 1; Length 540;
Best Local Similarity 60.2%; Pred. No. 2.4e-17;
Matches 260; Conservative 0; Mismatches 167; Indels 5; Gaps 3;

```

```

QY 859 tgcgcctagtgtgacacatccacagacctccaccaccatcaacaacatcttcgctgca 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 TGAAGCTACTGTGAGCATGTGCTGAAGAGATTAATAATTTTAACTCTTTGAGTTA 134
QY 919 tcgaagccgcctcagagccagatcactacgtgtcatcggggcccaaggaggtgcatgg 978
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 TTAAGGCTTTGTAGAACACAGATCACTATGTGTGATGGGCCAGAGAGATGCATGG 194
QY 979 ccccagagcagcctaatic-gcgtctgggagcgcgtactactctctgagctggtgcgacc 1037
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 195 GCCCTGGAGCTGCACAATCNCGTGTAGGACAGCTCTCTATTGAACCTTGCCAGATG 254
QY 1038 ttcttcctcatgg---tgagcaagcgcttcggcccgccgacagaagctcctctccatcagc 1094
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 TTCTCAGATATGCTTTAAAGATGGGTTTCAGCCCAAGAGCATTAATCTTTGCCAGT 314
QY 1095 tggagcgtgtgacttggaaagcgtctccacgagagtggtgtagaaggtacc-tcag 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 315 TGAAGTGTGAGAGACTTTGATGCGTTGCTGCACATGATGCGTAGAGGATACCTTTCC 374
QY 1154 cgtctgcacctcaagccgtagtgcagtgagccttgagacaagcagtgctggggatga 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 TONCCTGCATTAAAGGCTTTCACCTATATTAATTCGTGATTAACCGGCTTCTTGATCCAG 434
QY 1214 caaatttcacgaagaccacccctctcagaaagtcatagtctatagagtgcccgagca 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 CAACTTCAAGGTTTCTGCGCACCCACCTGTTGTATACGCTTATTGAGAAACATGCAAAA 494
QY 1274 gttgattctcc 1285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 495 TGTGAGCATCC 506

```

```

RESULT 3
T19421
ID T19421 standard; cDNA to mRNA; 93 BP.
AC T19421;
DT 04-JUL-1996 (first entry)
DE Human gene signature HUMGS00446.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN WO9514772-A1.
PD 01-JUN-1995.
PR 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR MPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1: Page 380; 2245pp; Japanese.
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS for
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
SQ Sequence 93 BP; 28 A; 17 C; 24 G; 23 T;

```

```

Query Match 3.5%; Score 89.4; DB 1; Length 93;
Best Local Similarity 97.8%; Pred. No. 4.3e-10;
Matches 90; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2421 gatcgtggcccatagacttatggccacagcgtgctgtggtgaaagggcggtgagt 2480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GATCGTTGGCACATAGCTTATGCGCACACAGGTGCTGTGTGTGTAAGAGGGCGGTGAGT 60
QY 2481 ttcaatatcaataaacaacctgatatacaata 2512
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 TTCAATATCAATTAACACACCTGATATCAATTA 92
RESULT 4
Q65336
ID Q65336 standard; cDNA; 660 BP.
AC Q65336;

```

DI 11-JAN-1995 (first entry)
DE Sequence used in identification of PSM Ag cDNA.
KW Prostate-specific membrane antigen; PSM; prostate cancer;
KW transmembrane glycoprotein; imaging; targeting; tumour detection;
KW antibody detection; ds.
OS Homo sapiens.
PN W09409820-A.
PD 11-MAY-1994.
PI 05-NOV-1993; US-973337.
PR (SLOK) SLOAN KETTERING INST CANCER.
PA Fair WR, Heston MDW, Israeli RS;
PI WPI; 94-1617129/20.
PT Prostate-specific membrane antigen and DNA encoding it - is
PT useful for detecting haematogenous micro-metastatic tumour cells
PT and for identifying ligands which bind to PSM Ag
PS disclosure; Page 120; 156pp; English
CC This sequence is given in the specification but is not mentioned in
CC the disclosure. The PSM coding sequence is useful for suppressing or
CC modulating the metastatic ability of prostate tumour cells to grow,
CC or for eliminating them. The protein is useful to identify or purify
CC ligands of the Ag. It is also an attractive target for Ab-directed
CC imaging and targeting of prostatic tumour deposits.
SQ Sequence 660 Bp; 190 A; 126 C; 147 G; 181 T;

Query Match	3.5%	Score 87.2	DB 1	Length 660
Best Local Similarity	55.8%	Pred. No.1 6e-01		
Matches 240	Conservative	0	MissMatches 181	Indels 9
				Gaps 4
QY	859	tgcgcctagtggtcacaacatcacagagactccacccccatacaacatcttcgtgtgca	918	
Db	96	TCAGATCAGCTGTGAACATGTACTGTAAAGAAACAGAAATTAATTAACATCTTTGGCGTTA	155	
QY	919	tctgaagccgcgtcccaagaccagatacactgtgtcactcgtgggcccagagaggtatgcat	978	
Db	156	TTAAAGGCTATGAGGAACCGACCCGTACATTTGTATAGAGCCCGAGAGAGCGCTTGGG	215	
QY	979	ccccagg-aggacgataatccgctgtgtgggagc-ggctatactcctgagctgtgtggac	1036	
Db	216	GCCCTGTGTTGGTGGCAAGTCAGATGTGGGGAACAGGCTTTCCTTTAAACTTCCCAAGT	275	
QY	1037	ctttccctccatggtgtgagcaac---ggcttcggcccccgagaagctccctcttcacag	1093	
Db	276	ATTCTCAGATGATGATTTCAAAGATGATGATTTAGACCCACAGAGATTAATCTTTGCCAG	335	
QY	1094	ctggagacgctgtgtacttcttggaagcgttgggtctccaaagtggtctgaaggtctactcag	1153	
Db	336	CTGGACTGCAGAGACTATGAGCTTTGGTCCGACTGTAGTGGCTGAGGGGTAACCTTTC	395	
QY	1154	cgctgcacacctcaaacgcg---tagtctagctgagcccttgcgaacgcagctgtcgggg	1209	
Db	396	ATCTTGCACTCAAAAGNNNGCTTTCCTACTTAATTAATNCTGATAAAGTCGCTCGGTA	455	
QY	1210	atcgacagttctatgcgaagaccagagccccctctctgcacagctcattcattgagatgtctc	1269	
Db	456	CTAGCACTTCAAGGTTTCTGCCAGCCCCCTATTATATACACTTATGAGGGAAGATATGC	515	
QY	1270	agcaggtgga 1279		
Db	516	AGGANNCGTA 525		
RESULT	5			
ID	065520	standard; cDNA; 2653 BP.		
AC	065520:			
DT	11-JAN-1995	(first entry)		
DE		Prostate-specific membrane antigen cDNA.		
KW		Prostate-specific membrane antigen; PSM; prostate cancer;		
KW		transmembrane glycoprotein; Imaging; targeting; tumour detection;		
OS		antibody detection; ds.		
OS		Homo sapiens.		

EH	Key	Location/Qualifiers
FI	Cds	262..2514
FT		/*tag= a
FT		/product= prostate specific membrane antigen (PSM)
PN	W09409820-A.	
PD	11-MAY-1994.	
PF	05-NOV-1993; U10624.	
PR	05-NOV-1992; US-973337.	
PA	(SLOK) SLOAN KETTERING INST CANCER.	
PI	Fair WR, Heston MDM, Israeli RS;	
DR	WPI: 94-167129/20.	
DR	P-PSDB: R55097.	
PT	Prostate-specific membrane antigen and DNA encoding it - is	
PI	useful for detecting haematogenous micro-metastatic tumour cells	
PI	and for identifying ligands which bind to PSM Ag	
PS	Claim 3: Page 103-106: 196pp: English	
CC	O65520 encodes a prostate specific membrane antigen (PSM, R55097).	
CC	The PSM coding sequence is useful for suppressing or modulating the	
CC	metastatic ability of prostate tumour cells to grow, or for	
CC	eliminating them. The protein is useful to identify or purify ligands	
CC	of the Ag. It is also an attractive target for Ab-directed imaging	
CC	and targeting of prostatic tumour deposits.	
SQ	Sequence 2653 BP: 762 A; 524 C; 640 G; 707 T;	

Query Match	2.1%	Score 53.8	DB 1	Length 2653
Best Local Similarity	48.18%	Pred. No. 0.013		
Matches 145	Conservative	0	Mismatches 152	Indels 0
				Gaps 0
Qy	897	atcaacaacatcttcgtgcgcgcgcgaagccgctcagaagccgagatcaactacgtgtgcac	956	
Db	1324	ATTTCACATNGTATATAGTACTCTCAGAGGAGCAGTGGAAACCGACAGATATCTCATTTCTG	1383	
Qy	957	ggggcccaaggagatgatgggccccaggagcagctaaatccgctgttgggagcgactata	1016	
Db	1384	GGAGGTACACGGGAGCTCATATGGCTGTTGGTGGATTGACCTTCAGAGTGGAGCACCTGTT	1443	
Qy	1017	ccccggagctgtgtgcgaaccttcccccacatggtgagcaagcgtccgcgcgcgaca	1076	
Db	1444	GTTTCATGAAATGTGTAGGAGGCTTTGGACACTGAAGAAAGGAGGTGGAGACTTGAAAGA	1503	
Qy	1077	agtcctcttcacatcagctgcgagacggtgtgtgactttgaaagcgltgggtccacgga	1136	
Db	1504	ACAATTTTGTTCGAAGCTGGGATGTCAGAGAAATTTGCTCTTGCTTCTACTGAGTGG	1563	
Qy	1137	ctagaagctacacctagcgtgtgcacctcaaacgcgtagtgttacgttgaagcctggac	1193	
Db	1564	GCAGAGGAGAAATTCAGAACTCCTTCAGAGGCTGGCGCTTATATTATGCTGTAC	1620	
RESULT 6				
ID	T36785	standard; cDNA; 2654 BP.		
AC	T36785;			
DR	04-NOV-1996	(first entry)		
DE	Prostate-specific membrane antigen cDNA.			
FW	Prostate-specific membrane antigen; PSM; promoter; prostate cancer;			
KW	metastasis; gene therapy; diagnosis; ss.			
OS	Homo sapiens.			
FT	Key	Location/Qualifiers		
FT	5'utr	1..261		
FT	cds	/*tag= a		
FT		262..2253		
FT		/*tag= b		
FT	misc_feature	/product= PSM antigen		
FT		114..380		
FT		/*tag= c		
FT		/note= "bases 114-380 (-147 to +109) are absent		
FT		in PSM' cDNA"		
FT		2352..2357		
FT	polya_signal	/*tag= d		
PN	WO9626272-A1.			
PD	29-AUG-1996.			

Query Match	2.1%	Score 53.8	DB 1	Length 2654
Best Local Similarity	48.8%	Pred. No. 0.013		
Matches 145, Conservative	0	Mismatches 152	Indels 0	Gaps 0

```

RESULT      7
ID          N70566
AC          N70566;
DC          26-APR-1991 (first entry)
DE          Sequence encoding functional part of corn phosphoenolpyruvate
DE          carboxylase (PEPCase).
KW          Enzyme; CO2 fixation; corn starch; photosynthesis; ss.
OS          Corn.
FH          Key
FT          cds
FT          Location/Qualifiers
FT          3..2810
FT          /*tag= a
          EP-212649-A.
          PN          04-MAR-1987.
          PD          22-AUG-1986; 111680.
          PR          23-AUG-1985; JP-186181.
          PA          (SUMO) SUMITOMO CHEM IND KK.
          PA          (KATSU/) KATSUKI H.
          PI          Katsuki H.
          DR          WPI: 87-056191/09.
          DR          P-PDB: P70362.
          New dna sequence coding for corn

```

Query Match	1.9%	Score	47.6	DB	1	Length	3059
Best Local Similarity	49.2%	Pred.	No. 0.24				
Matches 187; Conservative	0	Mismatches	184	Indels	9	Gaps	2

RESULT 8
V99230
ID V99230 standard; DNA: 1030 BP.
AC V99230;
DT 09-MAR-1999 (first entry)
DE DNA encoding an active acyltransferase domain (venA) from a PKS.
KW Enzymatically active acyltransferase domain; polyketide synthase; PKS;
KW polyketide synthesis; antimicrobial; anticancer; antifungal;
KW immunosuppressant; antihelmintic; venA; malonate acyltransferase domain;
KW PKS cluster; Streptomyces venezuelae ATCC15439; ds.

PN WO9851695-A2.
PD 19-NOV-1998.
PE 13-MAY-1998; U09518.
PR 16-MAY-1997; US-858003.
PA (ABSO) ABBOTT LAB.
PI Kakavas SJ, Katz L, Pereda-Lopez A, Ruan X, Stassi DL,
PI Summers RG.
DR WPI; 99-070114/06.
DR P-PSDB; W8713.
PT New polyketides - produced by microorganism having altered polyketide
PT synthase genes - especially chimeras containing a heterologous
PT acyltransferase domain, useful as e.g. antimicrobials, anticancer
PI agents
PS Claim 24; Fig 13; 137pp: English.
CC The present sequence encodes an enzymatically active acyltransferase
CC domain, venAT, from a polyketide synthase (PKS). The venAT protein
CC is derived from the malonate acyltransferase domain from the
CC PKS cluster of *Streptomyces venezuelae* ATCC15439. The sequence

```

RESULT      10
ID          107530
AC          107530;
DT          09-NOV-1998 (first entry)
DE          Human transcription factor YY1 CDNA.
KW          YY1; transcription factor; human immunodeficiency virus; HIV; AIDS;
            infection; retrovirus; therapy; LSF; ss.
OS          Homo sapiens.
FH          Key
FT          location/Qualifiers
            241..1485
            CDS
            /*tag= a
PN          W09833067-A1.
PD          30-JUL-1998.
PF          13-JAN-1998; U00574.
PR          23-JAN-1997; US-036242.
PA          (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.
PI          Devico A, Margolis D, Romero F;
            WPI; 98-428099/36.
DR          P-PSDB: W65406.
DT          Transcription factors YY1 and LSF, and their derivative(s) and
            analogue(s) - useful for. e.g. inhibiting HIV transcription.
PT          replication and/or infection in vitro or in vivo or preventing
            disorders associated with HIV infection.
PS          Disclosure; Page 74-76; 112pp; English.
CC          This is the nucleotide sequence of cDNA encoding human YY1 (see
            W65406), a multifunctional transcription factor that has previously

```


OY	1879	gagaccacacgtggcgccgcctcgcgcgcacactcgtgctgcttgcctccaacatctcg	1938
Db	28097	TGACCTTCGCCGAAGTACGGCTTTGCACCACCGCCTACTACTGTGGCGCCGCCGCAATCCG	28156
OY	1939	ggaccgccgggagc	1951
Db	28157	CGACCGACGCGGC	28169
RESULT	12		
ID	Q37543/c		
AC	Q37543	standard; cDNA; 4356 BP.	
DT	17-JUN-1993	(first entry)	
DE	Cardiac adenylyl cyclase type V gene.		
KW	CACV; therapy; diagnostic; cardiac function; cyclic AMP; cAMP; heart;		
XN	failure; ss.		
OS	Canis familiaris.		
FH	Key	Location/Qualifiers	
FT	cds	148..3702	
FT		/tag= a	
PN	EP-529622-A.		
PD	03-MAR-1993.		
PF	27-AUG-1992.		
PR	29-AUG-1991; US-751460.		
PA	(AMCY) AMERICAN CYANAMID CO.		
PI	Ishikawa Y, Koniski AF;		
DR	WPI; 93-06868/09.		
P-PSDB:	R32882.		
PT	Isolated nucleic acid mol. encoding Cardiac adenylyl cyclase type		
PS	V - useful for determining and modifying cardiac function		
CS	Claim 1; Page 15-27; 38pp; English.		
CC	Left ventricular tissue of canine heart was used as a source of mRNA.		
CC	A CDNA library was prepd. in lambda gt10 phage. A 970 bp Aat-HincII		
CC	fragment from type I adenylyl cyclase cDNA was used as probe. The		
CC	clones isolated were used to obtain cDNA encoding CACV. This probe		
CC	may also be used to screen a human cardiac cDNA library to obtain		
CC	the cDNA encoding human CACV. CACV, its analogues and antibodies		
CC	are useful in therapy or diagnostic assays, e.g. in modifying and		
CC	determining cardiac function. A decrease in CACV content of the		
CC	heart contributes to impaired CAMP prodn. and in heart failure. The		
CC	CACV can also be used to screen for cpds. which stimulate or inhibit		
CC	the activity of the cyclase.		
SO	Sequence	4356 BP; 837 A; 1367 C; 1377 G; 775 T;	
Query Match	1.8%; Score 45; DB 1; Length 4356;		
Best Local Similarity	49.4%; Pred.No. 0.86;		
Matches	117; Conservative 0; Mismatches 120; Indels 0; Gaps 0;		
OY	1506	ggccgcctgcgcgcgcgttgccaggcgcgtgtgcgcacagtctgcgcagtcgaactctcatccgg	1565
Db	469	GCCCTTGCGGCTCGGGCTCGGCGCGCCGCCGCCGCCCTCGTCCAGGCCAGCTCCAC	410
OY	1566	ctcaagcacagatcgcctgtcgtccctcgcacttcggcgcgcgtacaggagagctgtctcaag	1625
Db	409	CGAGCGGGGGCGCACACTCGCCCCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGACG	350
OY	1626	caaatcgggaacctcaagagttctcttgggaacctcaagcccgcgcgtgtgacctgcaag	1685
Db	349	GCGCGCCCGGAGACTCGCCCCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCG	290
OY	1686	tgggtgtactcggcgcgggggagactacatcgcggcgcggaagaactcgtcgcgagag	1742
Db	289	CAGTCGTGCGCCGCCGCTCTGCGAGCGCCGACCTGGAACGGAACCTGAACGGAAG	233
RESULT	13		
ID	Q95540/c		
AC	Q95540	standard; DNA; 4356 BP.	
DT	31-JAN-1996	(first entry)	

DE	Cardiac adenylyl cyclase gene.
KW	Cardiac adenylyl cyclase; effector enzyme; ss.
OS	Homo sapiens.
FT	key
FT	Location/Qualifiers
FT	148..3702
FT	/tag= a
PN	TW-243453-A.
PD	21-MAR-1995.
PF	02-JUL-1992; 105242.
PR	12-JUN-1992; US-899068.
PA	(AMCY) AMERICAN CYANAMID CO.
DR	WPI; 95-214006/28.
DR	P-PsDB; R78519.
PT	Cardiac adenylyl cyclase and corresp. DNA - having specified sequences
PS	Claim 1; Fig 2; 45pp; Chinese.
CC	Q95440 encodes R78519, the novel effector enzyme cardiac adenylyl cyclase.
SQ	Sequence 4356 BP; 836 A; 1372 C; 1373 G; 775 T;
Query Match	1.8%; Score 45; DB 1; Length 4356;
Best Local Similarity	49.4%; Pred. No. 0.86;
Matches 117; Conservative 0; Mismatches 120; Indels 0; Gaps 0;	
OY	1506 ggcgcctgcccgcgcgtggccagcgatgccagctcgcaggacgtccatccgg 1565
Db	469 ggccctggggctggggctggccggccggccggccggccggcgcgcgcacac 410
OY	1566 ctcaaccagatcgcctctgtgcacctgaatttcgcgcgtacggagtgactcagg 1625
Db	409 CGAGCGGGGCGCACCTCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCAGG 350
OY	1626 ccacatcggaacctcaacagatctcttggagacctcaagccgcgcggagtacctcag 1685
Db	349 GGCGCGCCGGGACTCCCCCGCCCGCCCGCCCGCCCGCCCGCCGTGCCTCGCCG 290
OY	1686 tgggtgtactcggcggggggagactacalcggcggcggaagaagctcggcaggag 1742
Db	289 CAGTGTGCGCGCCGCGCTCTCTCCAGCGCGACGTGAGCGAAGCTCAAGCCGAAG 233
RESULT 14	
V45824	
ID	V45824 standard; DNA; 6781 BP.
AC	V45824:
DT	09-NOV-1998 (first entry)
DE	MaiZe phenolpyruvate decarboxylase gene.
KW	Phenolpyruvate carboxylase; PEPc; maize; corn; C3 plant; rice;
KW	C4 plant; photosynthesis; transgenic plant; ss.
OS	Zea mays.
FH	Key
FT	location/Qualifiers
FT	1185..1189
FT	/tag= a
FT	CAAAT_signal
FT	842..846
FT	/tag= b
FT	protein_bind
FT	1128..1233
FT	/tag= c
FT	/note= "Sp1 binding site"
FT	1160..1165
FT	/tag= d
FT	protein_bind
FT	1488..1493
FT	/note= "Sp1 binding site"
FT	/tag= e
FT	protein_bind
FT	1494..1499
FT	/note= "Sp1 binding site"
FT	/tag= f
FT	misc_feature
FT	552..560
FT	/tag= g
FT	/function= light responsive element
FT	663..680
FT	/tag= h

RESULT	15
ID	X26879 standard; DNA; 2135 BP.
AC	X26879;
DT	23-JUN-1999 (first entry)
DE	DNA encoding a protein with cation transporting activity.
KW	Organic cation transporter; OCT1; OCT2; drug development; fatty liver; heart disease; cancer; anti-tumour drug; anticancer drug; ss.
OS	Homo sapiens.
FT	key
FT	Location/Qualifiers
FT	CDS 147..1802
FT	/*tag- a
PN	W09913072-A1.
PD	18-MAR-1999.
PE	07-SEP-1998; J04009.
PR	20-MAY-1998; JP-156660.
PR	08-SEP-1997; JP-260972.
PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PI	Nezu J, Oku A;
DR	WPI: 99-215062/18.
DR	P-PSDB: Y01649.
PT	Genes homologous with organic cation transporters OCT1 and OCT2, useful in design of new drugs for treatment of diseases due to abnormality of the transporter functions
PS	Claim 2; Page 45-51; 97pp; Japanese.
CC	The present sequence encodes a protein with cation transporting activity. The genes are significantly homologous with organic cation transporters OCT1 and OCT2. The genes may used in drug development, particularly in the treatment of diseases due to abnormality of the organic cation transporter functions e.g. fatty liver, heart diseases and cancers, by controlling such as by inhibition or activation.
CC	Administration of anti-tumour and anticancer drugs in combination with a transporter protein inhibiting agent allows the agents to penetrate into the diseased cells to enhance the drug action.
SO	Sequence 2135 BP; 499 A; 530 G; 559 T;

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 13, 2000, 11:10:38 : Search time 1456.15 Seconds
(without alignments)
5929.991 Million cell updates/sec

Title: US-09-358-755-3

Perfect score: 2519

Sequence: 1 ggcgcgcgggagcgctct.....ctgatatcaataagccaana 2519

Scoring table: IDENTITY_NUC

Searched: 4089388 seqs, 171395092 residues

Database: EST.*

Word size: 0

Number of hits that pass the threshold : 8178776

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*

50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_gss1:*
70: gb_gss2:*
71: gb_gss3:*
72: gb_gss4:*
73: em_gss1:*
74: em_gss2:*
75: em_gss3:*
76: em_gss4:*
77: gb_gss5:*
78: gb_gss6:*
79: gb_gss7:*
80: gb_gss8:*
81: gb_gss9:*
82: em_gss5:*
83: em_gss6:*
84: em_gss7:*
85: em_gss8:*
86: em_gss9:*
87: em_gss10:*
88: em_gss11:*
89: gb_gss10:*
90: gb_gss11:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	544.4	21.6	548	61	AI814648	AI814648 wj75f09.x
2	521.6	20.7	536	38	AA748421	AA748421 ny01604.s
3	481.4	19.1	700	50	AJ225643	AJ225643 RNJ225643
4	443.6	17.6	504	27	AA007639	AA007639 zh99408.r
5	414.4	16.5	782	43	AU035332	AU035332 AU035332
6	394.8	15.7	452	23	R99295	R99295 yq71d03.s1
7	389.6	15.5	444	22	R46386	R46386 yq50d07.s1
8	383.6	15.2	425	22	R49459	R49459 yq58f02.s1
9	371.8	14.8	444	27	AA007614	AA007614 zh99408.s
10	367.2	14.6	617	41	AI047358	AI047358 ud65d02.y
11	363	14.4	384	34	AA505641	AA505641 nh82d03.s
12	351	13.9	388	37	AA682374	AA682374 zj86d08.s
13	350.2	13.9	363	38	AA809686	AA809686 zj86d08.s
14	334.6	13.3	384	24	H63101	AA809686 zj86d08.s
15	334	13.3	366	24	H70432	H63101 yf50h04.s1
16	324.8	12.9	503	22	R35943	H70432 yf50h04.s1
17	318.2	12.6	341	37	AA693893	R35943 yf50h04.s1
18	299.6	11.9	367	26	W89178	AA693893 zj49d07.s
19	292.6	11.6	708	41	AI046365	W89178 zh74h11.s1
20	267.4	10.6	408	21	T64983	AI046365 ud65d02.x
21	266.2	10.6	333	21	T98167	T64983 yd11e08.s1
22	236.2	9.4	421	45	AI386238	T98167 yf56c10.s1
						AI386238 mm20c12.y

ALIGNMENTS

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmett-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/dbp/image/image.html

```

"organism":"Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2408681"
/clone_1b="NCI-CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"

```

	/lab host="DH10B (phage-resistant)"				
	/note="Organism: lung; Vector: p7F73-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7F73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bernaldo."				
BASE COUNT	126 a	135 c	165 g	120 t	2 others
ORIGIN					
Query Match		21.6%;	Score 544.4;	DB 61;	Length 548;
Best Local Similarity		99.5%;	Pred. No. 5.3e-99;		
Matches 545;	Conservative	0;	Mismatches	3;	Indels
				Gaps	0.
Oy	1968 ttccaggggaaacgcgttttcaggcgctcaagttagcgccttgaaccgttgtaaggaga	2027 			
Dd	548 TTCGAGGAGACCGCTTCGGGTGAAGTAGCCCTGCATCCTTGAGCAAGGGAGA	489 			
Yl	2028 gccaatggacttagcgggagatgctctgacaatttgtataaacaaaccttgtagcccccttgggat	2087 			

QY	2088	cctcaacatcccgcctcccccagctcaagaagctcctcgtcctcgtcttgaaatgattcaaggt	2147
Db	428	CCTCAcATccCGcTccCCcAGTcAAAGcTccCTcTcTcTcGTGAAATGATTAAgGgT	369
QY	2148	caaggagatgagtcacagatlcacacttcacatgctgatacaatttcacatbcccctacaca	2207
Db	368	CAGGAGAGTGGCTCAAGATCCACCTCTCATCTGTGATCAATTTCTCATTTACCCCTTACACA	309
QY	2208	ttcttcacagagacccaagaccccaagacagatatcacacaccccaagccctgcagttgag	2265
Db	308	TCTCTCCAGGAGACCCAGACCCAGACAGATATCCACACACCCAGCCCTGCAATGTAG	249
QY	2268	ctgacccctaatgtgacaggtcatactgcttgattaatcagaagtagcatccctccatcac	2327
Db	248	TTGACMTAATGTGAGGGGTATCTGTGCGTTATATCAGAGATAGATCCCTTCAATCAC	189
QY	2328	agcccttcccttctctgagggtcttcacatacctagagacacactctgggaagttctgtag	2387
Db	188	AGCCCCCTTCCCTTCTGTGGGTCTCTCAATACCTAGAGACACCTGTGGGAGGTTTCTTAGG	129
QY	2388	ccctgggaacctgacagctctgttagtgggaagaatcgcctggcacatagaaccttaggcc	2447
Db	128	CCCTGGGACCTGGCCAGCTCTGTTAATGGGAGAGATCGCTGGACATTAAGCTTAAGGCC	69
QY	2448	aacagatgagctctggatgaaagggcgctggagtttcaatatacaataaacacactgatac	2507
Db	68	AACAGAGTGCTGTGGTGAAGGGGCGTGGAGTTTCAATATCAATATTAACACACCTGATATC	9
QY	2508	aataagcc 2515	
Db	8	AATTAAGCC 1	

RESULT	2
LOCUS	AA748421/C
DEFINITION	AA748421 536 bp mRNA EST 18-FEB-1996 ny01a04.s1 NCI-GAP GCBI Homo sapiens CDNA clone IMAGE:1270404 similar to SW-TRSR_HUMAN P02786 TRANSFERRIN RECEPTOR PROTEIN mRNA sequence.
ACCESSION	AA748421
NID	92788379
VERSION	AA748421.1 GI:2788379
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia
	Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 536)
 NCBI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1798396.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCBI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www.bio.lnl.gov/bdrip/image/image.html

FEATURES
 source
 Insert Length: 1516 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 385.
 Location/Qualifiers
 1..536
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="1"
 /clone="IMAGE:1270446"
 /clone_1db="NCI CGAP.GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: p773D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20⁺, IgD⁺).
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CEBR). CDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-TGTACCAATCTGACAGTGGAGGCGGCTCATTTTCTTTTCTTTT-
 3']. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p773D vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 119 a 125 c 160 g 132 t
 ORIGIN

Query Match 20.7%; Score 521.6; DB 38; Length 536;
 Best Local Similarity 99.2%; Pred. No. 1.8e-94;
 Matches 524; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1992 cagctagccctgcactgcagctgcaaggcgacgcaatgcttagcgaggatgctc 2051
 Db 536 CAGCTAGCCCTGCTOACTGAGCCTGCAAGGCGGCAAGCGCTTAGCGGAGATGC 477
 QY 2052 tggacaattgtaacaactctgagggccttgaggatcccatcccccagctca 2111
 Db 476 TGAACAATGTATACAACTTGTAGGCGGCTCTCATACATCCCGTCCCAAGCA 417
 QY 2112 agagctcctgcctcctgcctgctgaattgattcaaggtaaggatggtcagatccac 2171
 Db 416 AGAGCTCTCTGCTCCCTGCTTGAATGATTCAGGGTGCAGGAGTGCGTCAAGATCCAC 357
 QY 2172 tctatgctgataatcttcattaccctcacacatctctcagaggagcccaacccca 2231
 Db 356 TCTATGCTGATCAATTCATTAACCCCTACACATCTCTCCAGGAGCCCAAGCCCA 297
 QY 2232 gcacagatatcacacacccagcctgcaagttagctgacccaatgtgaaggtcatatc 2291
 Db 296 GCACAGATATCTCACACACCCAGCCCTGCAGTGTAGCTGACCTTAATGTAGCGGTCA 237

QY 2292 tctcgttaatcacagagatgagatcccttcaatcacaccccttcccttttgggtcc 2351
 Db 236 TCTCGTTAATCACAGAGATGACATCCCTTCATATACAGCCCTTCTTGGGGTCC 177
 QY 2352 tccataccagaacacactctggaaggcttctgtagagccttggagcctgcagctctgt 2411
 Db 176 TCCATACCTAGAGACCACTCTGGGAGGTTGCTTAGGCCCTTGAGACCTGGCAGCTGTGT 117
 QY 2412 agtggagagatgcgcgcacacataagccttatggccaaaggtgtgtgtgaaagg 2471
 Db 116 AGTGGAGAGATGCTGCTGGACCATAGCCTTATGGCCAAACAGGTGTGTGTGTAAGGG 57
 QY 2472 gctggtggttcaatcatcaataaacccctgatatcaataagcccaaa 2519
 Db 56 GCGTGGAGTTTCATATCAATTAACCACTGATATCAATAAAAAA 9

RESULT 3
 AJ225643 700 bp mRNA EST 14-MAY-1999
 LOCUS RNU225643 Rat liver ESTs (E.Olivier) Rattus norvegicus cDNA clone
 DEFINITION IRL413, mRNA sequence.
 ACCESSION AJ225643
 NID 94833473
 VERSION AJ225643.2 GI:4833473
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 700)
 AUTHORS Olivier,E., Soury,E., Risler,J.L., Smith,F., Schneider,K.,
 Lochner,K., Jouzeau,J.Y., Fey,G., and Sallier,J.P.
 TITLE A novel set of hepatic mRNAs preferentially expressed during an
 acute inflammation in rat represents mostly intracellular proteins
 JOURNAL Genomics 57 (3), 352-364 (1999)
 MEDLINE 99263497
 COMMENT On May 6, 1998 this sequence version replaced gi:3114939.

CONTACT: E. Olivier
 U78 INSERM
 543 chemin de la Breteque, 76233 Bois-Guillaume, France
 Library construction: J.B.C. Vol.270, p2998-30006 (1995)
 POLY-A-No.

FEATURES
 source
 Location/Qualifiers
 1..700
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="IRL413"
 /clone_1db="Rat liver ESTs (E.Olivier)"
 /tissue_type="liver"
 /dev_stage="adult"
 /note="Organ: liver; Vector: lambda ZAPII; Library
 construction: Ripberger,J.A. et al. J. Biol. Chem. (1995)
 270(50):2998-30006"

BASE COUNT 149 a 198 c 210 g 143 t
 ORIGIN

Query Match 19.1%; Score 481.4; DB 50; Length 700;
 Best Local Similarity 84.8%; Pred. No. 2e-86;
 Matches 598; Conservative 0; Mismatches 101; Indels 6; Gaps 5;

QY 1106 tgactttggaagctggtgctccagaggtgctgaaggctcaactgagcgtgcacct 1165
 Db 1 TGACTTTGGCAGCGTGGAGCCACGGAGTTC-TGGAGGCTTACTCAGCGTGTACACT 59
 QY 1166 caaagcgttagtgaactgagcctggaacacgcagtgctggggagatgacaagtctatgc 1225
 Db 60 CAAAGCGTGTGCTGATGTAGAGCCGTGACAACTCGCTTGTGGGAACGGCAATTCATTC 119
 QY 1226 caagacagacccctcttgacaagctcattagagatgtcctgaagcaggttgatctcc 1285

Db	339	CCAGATATCCACACAGGAGGCTGTCTTAGGCCACACAGGCTGTGACGACATGTGCACCT	398
Oy	644	gggaactggagaacccctcaacaccccttggtcttccttccttcatacaaacaggttcctcc	703
Db	399	GGGAACCTGGAGACCCCTTACACACTGCTTCCCGTCCCTTCATCAACACCGATTCCTTC	458
Oy	704	agttgcatcattcagagcttcccaagatcccaagcccaatcagttgagacattgcttc	763
Db	459	AGTAGAATCATCAGGCGCTTCCAGCATCCCGCCGACCCATTCAGTGTGACATTCGTGA	518
Oy	764	ccgcctgctgaggaagctcaagaagccctgtgtgtcccccaagaatgtcaggaggctctc	823
Db	519	TCAATTGCTCAGGAACATCACAGGCGCCGCTGTGCTTCCAGAGANTGAAAGGACACCTMTN	578
Oy	824	agagccccccttcaaccctgggccccgggcccagacagctatgtgtgtcaacaatcacag	883
Db	579	AGGCTCTCTATTATTTGCTGGAGCTGGGCGCCGACTTACGCTTGAGGGGCAACACACCG	638
Oy	884	ga--cctccaccaccatcaacaacatcttcgctgctgacatcgaaagcc	927
Db	639	GAGTCTNTTACCCCATATGATAATTTTGGTGCATGANGGC	684
RESULT	6		
R99295/c	R99295	452 bp	EST
LOCUS	YGTJ003.s1	Soares fetal liver spleen	14-SEP-1995
DEFINITION	IMAGE:201221.3', mRNA sequence.		
ACCESSION	R99295		
NID	9985896		
VERSION	R99295.1	GI:985896	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 452) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,D., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R.		
TITLE	The Mashu-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	On Nov 29, 1993 this sequence version replaced gi:430583.		
FEATURES			
source	<p>Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Insert Size: 1674</p> <p>High quality sequence stops: 262 Source: IMAGE Consortium, LNLN This clone is available royalty-free through LNLN; contact the IMAGE Consortium (info@image.lnl.nyu.edu) for further information. Insert Length: 1674 Std Error: 0.00 Seq primer: Promega -21m13 High quality sequence stop: 262. Location/Qualifiers 1..452 /organism="Homo sapiens" /db_xref="GDB:370271" /db_xref="taxon:9606" /map="21" /clone="IMAGE:201221" /clone_lib="Soares fetal liver spleen INFLS" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)</p>		

R49459/c
 LOCUS R49459 425 bp mRNA EST 22-MAY-1995
 DEFINITION y68102.s1 Soares Infant Brain INIB Homo sapiens cDNA clone
 IMAGE:38253 3', mRNA sequence.
 ACCESSION R49459
 MID 9820357
 VERSION R49459.1 GI:820357
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 425)
 Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT On May 18, 1995 this sequence version replaced gi:811361.
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 1734
 High quality sequence stops: 326 Source: IMAGE Consortium, LNLN.
 This clone is available royalty-free through LNLN; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1734 Std Error: 0.00
 Seq primer: Promega -21ml3
 High quality sequence stop: 326.
 Location/Qualifiers
 1..425
 /organism="Homo sapiens"
 /db_xref="GDB:410794"
 /db_xref="taxon:9606"
 /clone="IMAGE:38253"
 /clone_lib="Soares Infant Brain INIB"
 /sex="female"
 /dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: lacmid BA; Site:1: Not
 I; Site:2: Hind III; 1st strand cDNA was primed with a Not
 I - oligo(dT) primer [5';
 AACTGAGAGATTCGCGCCGACAGAAATTTTCTTTTCTTTT 3'];
 double-stranded cDNA was ligated to Hind III adaptors
 (Pharmacia), digested with Not I and directionally cloned
 into the Not I and Hind III sites of the lacmid BA vector.
 Library went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."
 BASE COUNT 98 a 96 c 128 g 102 t 1 others
 ORIGIN
 Query Match 15.2%; Score 383.6; DB 22; Length 425;
 Best Local Similarity 96.9%; Pred. No. 5.2e-67;
 Matches 412; Conservative 0; Mismatches 10; Indels 3; Gaps 2;

OY 2274 ctaatgtacggtcactctgtcggtaatacagagatagcatccctcaatcacagccc 2333
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 245 CTAATGTACAGGTCATCTGTCGGTTATCATCAGAGATGACATCCCTCAATCACAGCCC 186
 OY 2334 ttcccttctgggggtctctcatccttagagaccactctgggaagttgtcaggccctcg 2393
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 185 TTCCCTTCTGGGGGTCTCTCATCTAGAGACCATCTGGGAGGTTGTAGGCCCTCGG 126
 OY 2394 gacctggccagctctgttagtggagagatcgtcgtgcacatacctatgtgccaacag 2453
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 125 GACCTGGCCACCTCTGTAGTGGAGAGATCGCTGGCACCATTAACCTTATGCCACAGG 66
 OY 2454 tggctctgtgtgaaggggcggtgagtttcaataataaaccactgtatataaag 2513
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 65 TGCTGTGTGGTGAAGGGCGGTGGAGTTTCATATCAATAAACACCTGATATCAATAAA 6
 OY 2514 ccaaa 2518
 ||||
 Db 5 AAAAA 1
 RESULT 9
 AA007614/c
 LOCUS AA007614 444 bp mRNA EST 09-MAY-1997
 DEFINITION zn95d08.s1 Soares_fetal_liver-spleen_INFLS_S1 Homo sapiens cDNA
 clone IMAGE:429423 3' similar to PIR:A48592 A48592 transferrin
 receptor protein - Chinese hamster ;, mRNA sequence.
 ACCESSION AA007614
 MID 91463600
 VERSION AA007614.1 GI:1463600
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 444)
 Hallier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
 Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
 and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 COMMENT On Apr 14, 1993 this sequence version replaced gi:315367.
 TITLE
 JOURNAL
 MEDLINE
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNLN; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 652 Std Error: 0.00
 Seq primer: mob.REGA+ET
 High quality sequence stop: 337.
 Location/Qualifiers
 1..444
 /organism="Homo sapiens"
 /db_xref="GDB:1329192"
 /db_xref="taxon:9606"
 /clone="IMAGE:429423"
 /clone_lib="Soares_fetal_liver-spleen_INFLS_S1"
 /sex="male"
 /dev_stage="20 week post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
 with a modified polylinker; Site:1: Pac I; Site:2: Eco RI;
 This is a subcloned version of the original Soares fetal

Db 315 GCCTTTCGGGGGCCACCTGAGTCTCTTCATGAGAGATGATCGGTGTACCA 374
QY 1455 ttccgtcacacaaggagacacttatggaactctcataaggtctgtcaaggccgctg 1514
Db 375 TTCCTGTACACGAGAGAGACATATGAAATCTGCACAGAATGCTGAGAGTCCCTG 434
QY 1515 cccgcgtgtgcccagagccgtgcccagctgcagggcagctctcacttcgctcagccac 1574
Db 435 CCCGCCGTGTGTCCAGGACAGTGGCTGAGCTCGCGGCCACTCTCTCATYCCGACTGAGCCAC 494
QY 1575 gatcgctgtgccccctgacttcggtcgctacgaggagctgctccacagcacatcgag 1634
Db 495 GATCACCCTACTCGCGTACACTTGGCGCCCTATGAGAGCTGGTGTCTAGGACATCGGC 554
QY 1635 aacctcaagagattctctgtggagacctcaaggccgcggctgacctcagtgagtgatc 1694
Db 555 AACCCCAATGAGTTCCTGTGGGACCTCAAGAGCGCGGCTGACCTGACATGCTGTGTAC 614
QY 1695 tc 1696
Db 615 TC 616

RESULT 11
AA505641/c 384 bp mRNA EST 18-AUG-1997
LOCUS nh82a03.s1 NCI_CGAP_Br1.1 Homo sapiens CDNA clone IMAGE:964972 3',
DEFINITION mRNA sequence.
ACCESSION AA505641
NID 92241778
VERSION AA505641.1 GI:2241778
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802599.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 2525 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 256.
Location/Qualifiers
1. 384
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="964972"
/clone_lib="NCI_CGAP_Br1.1"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
/note="Vector: p7773D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from pooled bulk
breast tumor tissue, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified p7773

vector. Library is not normalized. (The normalized
version of this library is NCI_CGAP_Br2.1) Library was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 85 a 86 c 108 g 105 t
ORIGIN

Query Match 14.4%; Score 363; DB 34; Length 384;
Best Local Similarity 98.7%; Pred. No. 6,4e-63;
Matches 366; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2149 agggaggtgtgctcagagctccactctcactgtgtgataatttcataccctacacat 2208
Db 384 AGGAGAGTGCTCAAGATCCACTCTCATCTGTGATCAATTTCTATTACCTTACACT 325
QY 2209 ctctcacagagaccagaccacagacagatalccacacaccacagccctgagtgatgc 2268
Db 324 CTCCTCAGGAGCCAGACCCAGCAGCATATCCACACACCCAGCCCTGAGTATAC 265
QY 2269 tgacctaatgtgacggtcactctgtgttaatacagagatagcatccctcaatcaca 2328
Db 264 TGACCTTAATGTGACGGTCTACTGTGCTTATATGAGATGATCCTTCAATCACA 205
QY 2329 gcccttcccccttctgtgggtctcactacataccagagaccactgtggaggttgctaggg 2388
Db 204 GCCCTTCCCTTCTGTGGGCTCTCATACCTAGAGACCACTGTGGAGGTTTGCTAGGC 145
QY 2389 cctggagacctggccagctctgtttagtggaagatcgcgtgacacatagaccttaggcca 2448
Db 144 CCTGGGACCTGGCCAGCTCTGTAGTGGAGAGATGCGTGGACCATACCTATTAGGCCA 85
QY 2449 acaagtggtctgtgtgaaaggcgctggagttcaataataaaccactgatataca 2508
Db 84 ACAGGTGTGTCTGTGTGAAGAGGGCGGTGAGTTCAATTAACCACTGATATCA 25
QY 2509 ataagccaaa 2519
Db 24 ATTAATAAAAAA 14

RESULT 12
AA682374/c 388 bp mRNA EST 19-DEC-1997
LOCUS zj36a08.s1 Soares_fetal_liver.spleen_INTLS.S1 Homo sapiens CDNA
DEFINITION clone IMAGE:461750 3' similar to gb:U1507 TRANSFERIN RECEPTOR
PROTEIN (HUMAN); mRNA sequence.
ACCESSION AA682374
NID 92669655
VERSION AA682374.1 GI:2669655
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
AUTHORS Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Schellenberg, K., Stepien, M., Tan, F.,
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
TITLE Washu-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394174.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 380.

FEATURES
SOURCE

Location/Qualifiers
1. 388
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="8"
/clone="IMAGE:461750"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen 1NFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', ACTCGAAGATTATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 65 a 108 c 129 g 86 t

ORIGIN

Query Match 13.9%; Score 351; DB 37; Length 388;
Best Local Similarity 97.3%; Pred. No. 1.6e-60;
Matches 357; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

379 ccatcgcaacgtcacgagagctgtgtacgcccactacagcgccgagacgtgc 438
|||||
388 CCATCGCAACGTACCGGAGAGCTGTGTAGCCCCACTACGGGGCCGAGACCTGC 329
|||||
439 aggacctgagggcgagggcggtgatccagtggcgccgtctgtgtgctggggg 498
|||||
338 AGGACCTCGGGGCCAGGGGGCTGTGATCCAGTGGGCGCTGTGTGCGGGGGG 269
|||||
499 tgatagctgtcccgagaggtgtacacatgtctcagagcttcggggctcaagaagtgtcca 558
|||||
268 TGATACCTGTGCCCCAGAGGTGACCAATGCTCAGGACTTGGGGCTTAAGAGATGCTCA 209
|||||
559 tataccagagcgagcgagcttctccagagaccaccacaaagcctgtgccagcgagc 618
|||||
208 TATACCCAGAGCGACGAGCTTCTCCAGAGACCCACCAAGCAGCCTGTCCAGACCGC 149
|||||
619 aggcagctataggacatgtgacactgtggaacttgagacccctcaacactgtctccctt 678
|||||
148 AGGCGATGATGACATGTGACCTGGAGACTGGAGACCCCTACACACTGTGCTTCCCTT 89
|||||
679 ccttcacaaaccagcttccctcctcctgtgtcatcatcagagcctccagacatccagccg 738
|||||
88 CTTTCAATCAAAACCAAGTTCCCTCCAGTTGATATCAAGGCTTCCAGCATCCACGCGC 29
|||||
739 agcccat 745
|||
28 GAGCCCT 22

RESULT 13
AA809686 363 bp mRNA EST 18-FEB-1998
LOCUS AA809686/c n299d04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1303591 3',
DEFINITION mRNA sequence.
ACCESSION AA809686
NID 92879092
VERSION AA809686.1 GI:2879092
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 363)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE
JOURNAL
COMMENT

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2286784.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at:
www-bio.llnl.gov/bdrip/image/image.html

Insert Length: 1482 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 361.
Location/Qualifiers
1. 363

FEATURES
SOURCE

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="954B04: 954B04"
/clone="IMAGE:1303591"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-GTGTACCAATCGAAGTGGAGCGGCGCCGCTCATTTTCTTTTCTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 83 a 77 c 106 g 97 t

ORIGIN

Query Match 13.9%; Score 350.2; DB 38; Length 363;
Best Local Similarity 99.2%; Pred. No. 2.2e-60;
Matches 352; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

2165 gtccacctctcatgtcgtatcatcttctcattaccctacacatctctcagagagcca 2224
|||||
363 GTCCACCTCTCATGTGATCAATTTCTCATTCCTTACACATCTCTCCACGAGGCCA 304
|||||
2225 gaccgccagcagatcacacacacaccccgccctgcagctgtatgtgacctaatgtacg 2284
|||||
303 GACCCAGACAGATATCCACACACCCACCCCTGACGTAGCTGACCTTAAGTGACG 244
|||||
2285 gtcatctgtcgttatacagagagtagcatcccttcataacacagcccttccttcctt 2344
|||||
243 GTCTACTGTGCTTATTCAGAGAGTAGCATCCTTTAATCAGACGCCCTTCCCTTTCT 184
|||||
2345 ggggtctccatattcagagacacactctgggaggtgtgtcctagcctggagcctggcag 2404
|||||
183 GGGGTCTCCATCTAGTACAGACCACTCTGGAGAGTGTCTAGGCCCTGGAGACCTGGCC 124
|||||
2405 ctctgttagtgggagagatgctgtgacacatagccttatggccaacgggtgtgtgtgct 2464
|||||
123 CTCTGTAGTGGGAGATGCTGCGACCAATAGCTTATGGCCAAAGGTGTGTGTGTGT 64
|||||
2465 gaaaggcgctggaggttcataatcaataaaccccgatatacataagccaana 2519
|||||

```

Db      63  GAAAGGGCGTGGAGTTTCAATATCAATAAACACCCTGATATCATTAATAAAAAA 9
RESULT  14
LOCUS   H63101/c
DEFINITION
H63101
ACCESSION
H63101
MID     91017902
VERSION H63101.1
KEYWORDS GI:1017902
SOURCE  EST.
ORGANISM human.
human.sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euteria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 384)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Holtman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Woldmann, P. and
Wilson, R.
The Washu-Merck EST project
Unpublished (1995)
On May 8, 1995 this sequence version replaced gi:799827.

TITLE   JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 656
High quality sequence stops: 272
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 656 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 272.
Location/Qualifiers
1. 384
/organism="Homo sapiens"
/db_xref="GDB:3777890"
/db_xref="taxon:9606"
/clone="IMAGE:206759"
/clone_lib="Soares fetal liver spleen JNLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and spleen; Vector: p7n3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[3], AACGGAAGATTAATTAAGACATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p7n3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 66 a 104 c 126 g 78 t 10 others
ORIGIN

Query Match 13.3%; Score 334.6; DB 24; Length 384;
Best Local Similarity 94.5%; Pred. No. 2.9e-57;
Matches 362; Conservative 0; Mismatches 19; Indels 2; Gaps 2;

Oy      353  ccctaagctctctccctacagccatcggaacgtcacagggagagctgtgtacgc 412
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db      383  CCTCTANGTCTACTGNCCTTAAGGCGCANNATCGNNAACGTCA-GGAGAGNCCTG6T7TAGC 325

Oy      413  ccactacggcgccgccgaagaccttcagagacttcggcgccagggcgctggatccagtggg 472

```

Db	324	CNACTAGGGGNGGCCGACACACTTGCAGAGACCTGCGGGCCAGAGGGCGTGATCCAGTGGG	265
Qy	473	ccgcctctctctgtgtgcgcgtgggggtg-atcagcttgcgccagaaggtgaccaatgtctc	531
Db	264	CCGGCTCTCTCTGTGTGGCGCTGGGGGTGAATACANNTGCCCCAGAAGGTGACATGTCTC	205
Qy	532	aggaactcggggctcaagagagtctcatatataccagaagcagcagactctccaggacc	591
Db	204	AGGACTTCGGGGGCTCAAGAGAGTGTCTCATATATACCAAGACAGCAGANGAGCTTCTCCAGGACC	145
Qy	592	caccacaagccaagcctgttccagccagcagcagcagtcgtatgaaatgtgcacttgggaactg	651
Db	144	CACCAACAGCAAGGCTGTCCAGCCAGCAGCAGTGTATGTGACATGTGCNCCCTGGGAACGTG	85
Qy	652	gagacccctacaacacctcgggtcttccttccttcaataaaccagcttccttcagttgat	711
Db	84	GAGACCCCTTACACACCTTGCGTTCCTTCTTCAATCAAAACCAAGTTCCTCCAGTTGGAT	25
Qy	712	catcaggccttccagcatcca 734	
Db	24	CATCAGGCTTCCAGCATCCA 2	
RESULT	15		
LOCUS	H70432/c		
DEFINITION	H70432	366 bp mRNA	EST 24-OCT-1995
ACCESSION	H70432		
NID	g1040638		
VERSION	H70432.1	GI:1040638	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Ekura-Yota; Metzozoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 366) Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B., Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacey, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierrey-Meg, J., Trevasakis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.		
TITLE	Generation and analysis of 280,000 human expressed sequence tags		
JOURNAL	Genome Res. 6 (9), 807-828 (1996)		
MEDLINE	97044478		
COMMENT	On Apr 14, 1993 this sequence version replaced gi:837502.		
FEATURES			
SOURCE	Source		
	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: estewatson.wustl.edu Insert Size: 801 High quality sequence stops: 269 Source: IMAGE Consortium, LINTL This clone is available royalty-free through LINTL; contact the IMAGE Consortium (info@image.lintl.gov) for further information. Insert Length: 801 Std Error: 0.00 Seq primer: Promega -21ml3 High quality sequence stop: 269. Location/Qualifiers 1..366 /organism="Homo sapiens" /db_xref="GDB:3779022" /db_xref="taxon:9606" /map="11" /clone="IMAGE:213718" /clone_lib="Soares fetal liver spleen INFLS" /sex="male"		

```

/dev stage="20 week-post conception fetus"
/lab host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' ACTGGAAGAATTAATTAAGATCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      89 a      81 c      107 g      85 t      4 others
ORIGIN

```

```

Query Match      13.3%; Score 334; DB 24; Length 366;
Best Local Similarity 98.9%; Pred. No. 3.8e-57;
Matches 356; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

OY 2153 agtggctcaggtccacctctcatgtg-ctgataatttctcattaccctcacatctc 2211
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 366 AGGTGNTCAGAGTCACCTCTCATTTGCGTATTCATTTTCATTAACCCCTANACATCTC 307

OY 2212 tccacgagagcccaagaccagatatccacacacccagccctgagtgtagctga 2271
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 306 TCCAGGAGCCCGAGACCCAGACAGATATCCACACACCCCTGAGTGTAGCTGA 247

OY 2272 ccctaattgacgtcactactgtcggttaatacagagtagatccctcaatacaagcc 2331
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 246 CCTATATGTGAGGTTCATCTGCGGTTAATCAGAGAGATGATCCTTCATTCACAGCC 187

OY 2332 ccttcaccttctgggggtccatatacctagaaccactctggaggttctgtagccct 2391
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 186 CTTTCCCTTTCTGGGGTCTCTCATACCTAGACCACTCTGGAGGTTTGCTAGGCCCT 127

OY 2392 gggacctgagccagctctgttaagtggagagatcgctggaccatagccttaaggcca 2451
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 126 GGGACCTGGCCAGCTCTGTAGTGGAGAGATCGCTGGACCATTA-CCTTATGGCCACA 68

OY 2452 ggtggtctgtgtaaaaggggcgtggagtttcaataataaacaacctgatatcaata 2511
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 67 GGTGCTGTGTGTAAGGGCGGTGAGATTTCATATCAATAAACCACTGATATCAATA 8

```

Search completed: January 13, 2000, 11:10:52
 Job time: 9276 sec